BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images, please do not report the images to the Image Problem Mailbox.

STIC-Biotech/ChemLib

77212

From: [redacted] Davis, Minh-Tam
Sent: Friday, October 04, 2002 9:00 AM
To: STIC-Biotech/ChemLib
Subject: Search request for 09/631863

priority
8/99

Please search in commercial database and in issued patent files:
SEQ ID NO:2
Thank you.
MINH TAM DAVIS
ART UNIT 1642, ROOM 8A01, MB 8E12
305-2008

Point of Contact:
Toby Port
Technical Info. Specialist
CM1 6A04
703-308-3534

	TYPE OF SEARCH:	VENDOR/COST (where applic.)
Searcher: _____	NA Sequences: _____	STN: _____
Phone: _____	AA Sequences: _____	DIALOG: _____
Location: _____	Structures: _____	Questel/Orbit: _____
Date Picked Up: 10/7	Bibliographic: _____	DRLink: _____
Date Completed: 10/4	Litigation: _____	Lexis/Nexis: _____
Searcher Prep/Review: //	Full text: _____	Sequence Sys.: _____
Clerical: _____	Patent Family: _____	WWW/Internet: _____
Online time: _____	Other: _____	Other (specify): _____

GenCore version 4.5

OM protein - protein search, using sw model

Run on: October 4, 2002, 15:38:18 ; Search time 20.92 Seconds
(without alignments)
1841.865 Million cell updates/sec

Title: US-09-631-863A-2
Perfect score: 2139
Sequence: 1 MRNKRVLKTKKRRSGRGGQD.....DNCPAKASKSSPAGNSPAPL 401

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_71:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	720.5	33.7	232	2	JE0163	myelin expression
2	240.5	11.2	853	2	S60178	gag polyprotein ho
3	190	8.9	457	2	T18347	gag protein homolo
4	159.5	7.5	639	2	S23569	gag polyprotein ho
5	148	6.9	1494	2	T13798	hypothetical prote
6	137	6.4	349	2	T18349	probable gag prote
7	135	6.3	537	1	FOMVRV	gag polyprotein -
8	134.5	6.3	555	2	T30349	structural protein
9	134	6.3	1529	2	A59310	unconventional myo
10	131.5	6.1	538	2	S70394	gag polyprotein -
11	131.5	6.1	1230	2	T18256	probable serine/th
12	131.5	6.1	1230	2	T18259	serine/threonine p
13	130	6.1	334	2	T01815	hypothetical prote
14	130	6.1	537	1	FOMVMB	gag polyprotein -
15	129	6.0	1110	2	T19673	hypothetical prote
16	129	6.0	1188	2	T46608	zinc finger protei
17	125	5.8	915	2	T26695	hypothetical prote
18	125	5.8	1651	2	T14160	transmembrane rece
19	123.5	5.8	1182	2	T30189	myelin transcripti
20	123.5	5.8	1585	2	T31611	hypothetical prote
21	122.5	5.7	536	1	FOMVME	gag polyprotein -
22	122.5	5.7	538	2	S35474	gag polyprotein -
23	122.5	5.7	601	2	S33377	P63 protein - huma
24	122.5	5.7	1316	2	T00381	KIAA0633 protein -
25	121.5	5.7	1187	2	T46637	transcription fact
26	121.5	5.7	1585	2	T18274	1-phosphatidylinos
27	121	5.7	1612	2	T30805	dutl protein - mo
28	120	5.6	745	2	D96829	homeobox protein (
29	120	5.6	747	2	S71478	homeotic protein A
30	120	5.6	992	2	T46337	hypothetical prote
31	120	5.6	1905	2	T18267	multidrug resistan
32	120	5.6	1937	2	I38055	myosin heavy chain
33	119.5	5.6	1870	2	S37671	MHC class III hist
34	119.5	5.6	1872	2	S36152	MHC class III hist
35	119.5	5.6	2954	2	T14156	kinesin-related pr
36	119	5.6	320	2	AE2842	conserved hypothet
37	119	5.6	359	2	F97619	hypothetical prote
38	119	5.6	465	2	A02986	myosin alpha heavy
39	118.5	5.5	428	1	I36930	involucrin - white
40	118.5	5.5	7962	2	I38346	elastic titin - hu
41	118	5.5	2175	1	S03170	homeotic protein c
42	117	5.5	972	2	T49773	related to actin-i
43	116.5	5.4	249	2	A37280	C/EBP-related prot
44	116.5	5.4	708	2	I83196	NEDD-4 ORF - mouse
45	116.5	5.4	2142	2	B35098	MHC class III hist

ALIGNMENTS

RESULT 1
JE0163
myelin expression factor-3 - mouse
N;Alternate names: MyEF-3
C;Species: Mus musculus (house mouse)
C;Date: 03-Jul-1998 #sequence_revision 10-Jul-1998 #text_change 07-May-1999
C;Accession: JE0163
R;Steplewski, A.; Krynska, B.; Tretiakova, A.; Haas, S.; Khalili, K.; Amini, S.
Biochem. Biophys. Res. Commun. 243, 295-301, 1998
A;Title: MyEF-3, a developmentally controlled brain-derived nuclear protein which spe
A;Reference number: JE0163; MUID:98139908
A;Accession: JE0163
A;Molecule type: mRNA
A;Residues: 1-232 <STE>
A;Experimental source: brain
C;Keywords: phosphoprotein
F;40-60/Domain: transmembrane #status predicted <TMM>
F;31,175/Binding site: phosphate (Ser) (covalent) (by protein kinase C) #status predi
F;31,142,185,201/Binding site: phosphate (Ser) (covalent) (by casein kinase II) #stat

Query Match 33.7%; Score 720.5; DB 2; Length 232;
Best Local Similarity 61.6%; Pred. No. 1.6e-38;
Matches 149; Conservative 24; Mismatches 52; Indels 17; Gaps 5;

```
Qy    154 CPEDLPEKFDGNPDMLAPFMAQCQIFMEKSTRDFSVDRVRVCFVTSMMTGRA--ARWASA 211
          | ||||||||||||||| ||| |||:|||||||||||||:||||||||: |||      :|
Db      2 CLEDLPEKFDGNPDMLGPFMYQCQLFMEKSTRDFSVDRIRVCFVTSMLIGRAPLGYCQAA 61

Qy    212 KL----ERSHYLMHNYPAFMMEMKHVFEDPQRREVAKRKIRRLRQGMGSVIDYSNAFQMI 267
          |:    : | |     | ::        |      : : :    | | |:||||||||||
Db     62 KMYLPDAQLHCLYDGAEAVLVR-------PSESVSVQTQDQTSAPGPGPVVDYSNAFQMI 114

Qy    268 AQDLDWNEPALIDQYHEGLSDHIQEELSHLEVAKSLSALIGQCIHIERRLARAAAARKPR 327
          |||||| ||||:||: |||:  |: |||  |  |:|:|||  |||||||||| ||| ||
Db    115 AQDLDWTEPALMDQFQEGLNPDIRAELSRQEAPKTLAALITACIHIERRLARDAAA-KPD 173

Qy    328 SPPRALVLPHIASHHQVDPTEPVGGARMRLTQEEKERRRKLNLCLYCGTGGHYADNCPAK 387
            |||||:|    : | |||||||||||::||||||||:|||||| |||:|| ||||
Db    174 PSPRALVMP---PNSQTDPTEPVGGARMRLSKEEKERRRKMNLCLYCGNGGHFADTCPAK 230

Qy    388 AS 389
          ||
Db    231 AS 232
```

RESULT 2
S60178
gag polyprotein homolog - fungus (Fusarium oxysporum) retrotransposon skippy
C;Species: Fusarium oxysporum
C;Date: 15-Feb-1996 #sequence_revision 01-Mar-1996 #text_change 21-Jul-2000
C;Accession: S60178

R;Anaya, N.; Roncero, M.I.G.
Mol. Gen. Genet. 249, 637-647, 1995
A;Title: Skippy, a retrotransposon from the fungal plant pathogen Fusarium oxysporum.
A;Reference number: S60178; MUID:96132549
A;Accession: S60178
A;Molecule type: DNA
A;Residues: 1-853 <ANA>
A;Cross-references: EMBL:L34658; NID:g510695; PIDN:AAA88790.1; PID:g510696
A;Experimental source: retrotransposon skippy
C;Genetics:
A;Mobile element: retrotransposon skippy
C;Keywords: polyprotein

Query Match 11.2%; Score 240.5; DB 2; Length 853;
Best Local Similarity 21.2%; Pred. No. 1.2e-07;
Matches 85; Conservative 65; Mismatches 157; Indels 93; Gaps 14;

Qy 33 GRSPPTPTVTLGPDCPPPPPPPPPNNNNNNNSKHTGHKSACVPNMTERRRDEL---SEEI 89
Db 19 GHQPAAP-------ANPAPVNRPPTDQNMDDADDSSQSSD--DSEVERLREQLGNVTNEM 69

Qy 90 NNLREKV-------MKQSEENNNLQSQVQKLTEENTTLR---EQVEPTPEDEDDDIELRG 139
Db 70 NEMRQMLEEFTALQHQQNQSNNNTQQEMYNLASAANNGRDPGEVLKPSP---------- 118

Qy 140 AAAAAAPPPPIEEECPEDLPEKFDGNPDMLAPFMAQCQIFMEKSTRDFSVDRVRVCFVTS 199
Db 119 ----------PEYFDGTPSKLPTFLTQSRAFITYYPNQFRNDSAKVMYMAG 159

Qy 200 MMTGRAARWASAKLERSHYLMHNY---PAFMME----------MKHVFEDPQ 238
Db 160 RLIKTAAQWFQP-------IMNDYMTNPPYKLQPRTALLFGENGRHEMEEALKMAFGTID 212

Qy 239 RREVAKRKIRRLRQGMGSVIDYSNAFQMIAQDLDWNEPALIDQYHEGLSDHIQEELSHLE 298
Db 213 EKGQAERKIKTLKQ-TGSASTLGVEFLQLASKLPWDQDVLMSFFFDALKEQVQQELWEKD 271

Qy 299 VAKSLSALIGQCIHIERR----LARAAAARKPRSPPRALVLPHIASHHQVDPT--EPVGG 352
Db 272 RPRTLVEYINMAVKIDDRQFAWRTRNSRGNKGRQDNKPRYHANQGRTRQTDTSYGTEAGP 331

Qy 353 ARMRLTQEEKERRRKLNLCLYCGTGGHYADNC--PAKASK 390
Db 332 MAIGMTKRDKSK----VTCYNCGKKGHYERECKNPVKTNQ 367

RESULT 3
T18347
gag protein homolog, truncated - rice blast fungus magnaporthe gypsy retrotransposon
C;Species: Magnaporthe grisea (rice blast fungus)
C;Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 02-Sep-2000
C;Accession: T18347
R;Farman, M.L.; Tosa, Y.; Nitta, N.; Leong, S.A.
Mol. Gen. Genet. 251, 665-674, 1996
A;Title: MAGGY, a retrotransposon in the genome of the rice blast fungus Magnaporthe gri
A;Reference number: Z18882; MUID:96335141
A;Accession: T18347
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-457 <FAR>
A;Cross-references: EMBL:L35053; NID:g522300; PID:g522301; PIDN:AAA33419.1
C;Genetics:
A;Gene: gag
A;Mobile element: magnaporthe gypsy retrotransposon

Query Match 8.9%; Score 190; DB 2; Length 457;
Best Local Similarity 21.8%; Pred. No. 9e-05;
Matches 77; Conservative 54; Mismatches 156; Indels 66; Gaps 9;

Qy 78 TERRRDELSEEINNLR----------EKVMKQSEENNNLQSQVQKLTEENTTL 120
Db 70 TDRVRDKAQADILAMAMEKASAITSRDAAFAELEKTRSELKEVRTVTLPTVHITTPAPTT 129

Qy 121 REQVEPTPEDEDDDIELRGAAAAAAPPPPIE--------EECPEDLPEKFDGNPDMLAPF 172
Db 130 NTPVEP----------LMVTPMGTTPPPASEHPASARLSERLPD--PDKFTGARSDLRRF 177

Qy 173 MAQCQIFMEKSTRDFSVDRVRVCFVTSMMTGRAARWASAKLERSHYLMHNYPAFMMEMKH 232
Db 178 ATQIRGKMTSNKDRFPNPESRLIYIAGRLSGKAYNLILPKMVGGTPQFGDYTDLLQYLEE 237

Qy 233 VFEDPQRREVAKRKIRRLRQGMGSVIDYSNAFQMIAQDLDWNEPALIDQYHEGLSDHIQE 292
Db 238 AFGDPDHVQNAQNKLYALKQRNVDFAEYLSEFQRLSLEGEMPEDALPPLLFQGISEELQD 297

Qy 293 ELSH--------LEVAKSLSALIG------QCIHIERRLARAAAARKPRSP------PRA 332
Db 298 MLLHNPAPSRQYHEFTRHLQSLDNRYRQHQQYKNRQTRTPRAAAPPARAAPRTQPDIPRA 357

Qy 333 LVLPHIASHHQVDPTEPVGGARMRLTQEEKERRRKLNLCLYCGTGGHYADNCP 385
Db 358 APKPNA----ELPLNDP-----MDLSSQRRHNRKENMLCYRCGSQEHFVAKCP 401

RESULT 4
S23569
gag polyprotein homolog - fungus (Cladosporium fulvum) retrotransposon CfT-1
C;Species: Cladosporium fulvum
C;Date: 12-Mar-1993 #sequence_revision 12-Mar-1993 #text_change 21-Jul-2000
C;Accession: S23569; S19849
R;McHale, M.T.; Roberts, I.N.; Noble, S.M.; Beaumont, C.; Whitehead, M.P.; Seth, D.;
Mol. Gen. Genet. 233, 337-347, 1992
A;Title: CfT-I: an LTR-retrotransposon in Cladosporium fulvum, a fungal pathogen of t
A;Reference number: S23569; MUID:92318885
A;Accession: S23569
A;Molecule type: DNA
A;Residues: 1-639 <MCH>
A;Cross-references: EMBL:Z11866; NID:g2562; PIDN:CAA77890.1; PID:g2563
A;Note: the authors translated the codon ACC for residue 55 as Ala, CAA for residue 4
C;Genetics:
A;Mobile element: retrotransposon CfT-1
C;Superfamily: Cladosporium fulvum probable gag polyprotein
C;Keywords: polyprotein

Query Match 7.5%; Score 159.5; DB 2; Length 639;
Best Local Similarity 24.0%; Pred. No. 0.011;
Matches 65; Conservative 37; Mismatches 114; Indels 55; Gaps 9;

Qy 159 PEKFDGNPDMLAPFMAQCQIFMEKSTRDFSVDRVRVCFVTSMMTGRAARWASAKLERSHY 218
Db 19 PEFFYGDRVKFDTWVSQMDMYFLFNS---MTENLKPIFATTFLRGRAQHWVKPFLRK--Y 73

Qy 219 LMHN----------YPAFMMEMKHVFEDPQRREVAKRKIRRLRQGMGSVIDYSNAFQMI 267
Db 74 LDSNGEDNADGVFKSYNHLKHAMKSVFGVSNEIATAVRVIQHLTQ-KTSTAEYAAKFQEY 132

Qy 268 AQDLDWNEPALIDQYHEGLSDHIQEELSH-LEVAKSLSALIGQCIHIERRLARAAAAR-- 324
Db 133 AQLTDWDDEALQVMYRRGLKEHVKDELMRDGRKIDGLGDLVQVTIDLDDKLYERAMERRY 192

Qy 325 ----------KPRSPPRALVLPHIASHHQVDPTEPVGGARMRLT 358
Db 193 DSKVSGKAGYTPGYDNRNRGFNDNYNKPKDKPYYGPQP-----MELDVTEK--GRKIRNS 245

Qy 359 QEEKE--RRRKLNLCLYCGTGGHYADNCPAK 387
Db 246 KGNRRPPSSRETRTCYGCGKPGHIARDCRGK 276

RESULT 5
T13798
hypothetical protein - fruit fly (Drosophila melanogaster) retrotransposon mdg3
C;Species: Drosophila melanogaster

```
C;Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
C;Accession: T13798
R;Avedisov, S.
submitted to the EMBL Data Library, February 1996
A;Reference number: Z17761
A;Accession: T13798
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1494 <AVE>
A;Cross-references: EMBL:X95908; NID:e990667; PID:e223896; PIDN:CAA65152.1
C;Genetics:
A;Cross-references: FlyBase:FBgn0002698
A;Mobile element: retrotransposon mdg3


  Query Match             6.9%;  Score 148;  DB 2;  Length 1494;
  Best Local Similarity   21.3%;  Pred. No. 0.16;
  Matches   89;  Conservative   53;  Mismatches  155;  Indels  120;  Gaps   22;

Qy      39 PTVTLGPDCPP--PPPPPPPNNN----------------------NNNNSKHTGHKSACV 74
           ||   | ||||  |    || |:                      |   |  ||  |
Db      41 PTAVRG-DCPPEHPQKNAPPGNDIFSSLDFQNCEINTDHVSVNAMNRKESTETG--SERE 97

Qy      75 PNMTE--RRRDELSEE---INNLREKVMKQSEE-----NNNLQSQVQK-------LTEEN 117
            || |  : | ||:|    :|  |  :  | ::        : | :|         |:||
Db      98 TNMFELQQLRAELAEAKAMLNGTRSSLQFQEQQQPEQSKATVSSVIQTAQFTQAGATKEN 157

Qy     118 TTLREQVEPTPEDEDDDIELRGAAAAAAPPPPIEEECPEDLPEKFDGNPDMLAPFMAQCQ 177
           ||           |     :   | |            ::    :||         |:
Db     158 TTFHSPQRSNERAESQRFPVDALALA-----------KETITDYDGKT------CARAW 199

Qy     178 IFMEKS-TRDFSVD--RVRVCFVTSMMTGRAARWASAKLERSHYLMHNYPAFMME----- 229
           | : |:  | |::|   :|:  :|  : | |  |          :| :|| ::|
Db     200 ITVVKNIARTFNIDDNHLRILLITK-LKGNAQVW----------LHAHPARLIEPIDNL 247

Qy     230 ---MKHVF-EDPQRREVAKRKIRRLRQGMGSVIDYSNAFQMIAQDLDWNEPALIDQYHEG 285
              :   | |   : |: ::   |  :   :   | :    ::  :: ::  |:||  ||
Db     248 LDQLSLTFGEQSSKAEIRRKFESRKWKTEENFCSYYDEKMALSNGINIDDDELLDQMIEG 307

Qy     286 LSDHIQEELSHLEVAKSLSALIGQCIHIERRLARA-AAARKP--RSPPRALVLPHIASHH 342
           :   :|   :   :          ||       : || :  | |  | ||
Db     308 IP--LQNFRTQARI---------QCFSTPSEMLRAFSNIRLPARREPP------------ 344

Qy     343 QVDPTEPVGGARMRLTQEE-------KERRRKLNLCLYCGTGGHYADNCPAKASKSS 392
            | ||:     |              |: :|:   |  ||  ||    || : | ||
Db     345 -VQPTDYKDAIRCANCNSRGHKADICKKPKREPGSCYACGQLGHLVAQCPTRKSVSS 400


RESULT   6
T18349
probable gag protein - rice blast fungus gypsy retroelement
C;Species: Magnaporthe grisea (rice blast fungus)
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 04-Mar-2000
C;Accession: T18349
R;Dobinson, K.F.
submitted to the EMBL Data Library, September 1994
A;Description: Sequence of the grh retroelement.
A;Reference number: Z18883
A;Accession: T18349
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-349 <DOB>
A;Cross-references: EMBL:M77661; NID:g538065; PID:g538066; PIDN:AAA21441.1
C;Genetics:
A;Mobile element: gypsy retroelement
C;Superfamily: rice blast fungus gypsy retroelement probable gag protein


  Query Match             6.4%;  Score 137;  DB 2;  Length 349;
  Best Local Similarity   23.3%;  Pred. No. 0.14;
  Matches   80;  Conservative   41;  Mismatches  124;  Indels   98;  Gaps   16;

Qy     105 NLQSQVQKLTEENTTLREQVEPTPEDEDDDIELRGAAAAAAPP--PPIEEECPEDLPEKF 162
           :|| :|| |     |:    |              |  | | |   |: :      |  :
Db      27 DLQGRVQALQTGAPTVPAIAE-------------ALQATALPKRKPLRD------PPLY 66

Qy     163 DGNPDMLAPFMAQCQIFMEKSTRDFSVDRVRVCFVTSMMTGRAARWAS-----AKLERSH 217
           || |   | | |       |  ||         |:          ||       |: ||:
Db      67 DGVP---ASFTAWRCAMEYKLRRDAD-------FIGDHRDQYEYLWAGLETSVQKVVRSY 116

Qy     218 YLM------HNYPAFMMEMKHVFEDPQRREVAKRKIRRLRQGMG-SVIDYSNAFQ---MI 267
           | :      : |  |:  ::  ::|| :|  |  ::  |:   | |   :   |:
Db     117 YEVGGRDGAYRYTDFLDYLERTYDDPHKRAQALAELETLKMKPGQSFAQFIAIFERTLAT 176

Qy     268 AQDLDWNEPALIDQYHEGLSDHIQEE--------------LSHLEVAKSLSALIGQCIH 312
           |  | | :    :     :|  |:|                  : :||: | |     |
Db     177 AGGLAWADEVRTNFLRFRVSPRIREACVGRGMGDGTYLGAVAIYRQVAQDLEA-----IE 231

Qy     313 IERRLA--RAAAARKPRSPPRALVLP--HIASHHQVDPTEPVGGARMR---------- 356
           ::||    || ||  || ||:    |   :|:      : | |||| |
Db     232 LDRRFGPHRAGAATAPR-PPKDEDTPMTGVAAM----GSRPNGGARGRRRPGQTQPSDTN 286

Qy     357 ----------LTQEEKERRRKLNLCLYCGTGGHYADNCPAKAS 389
                    :  :| :|||:   || ||  ||   :|   |:
Db     287 RRDTRPRAQWVPSDEYQRRRETGACLRCGNSGHQVADCTYAAA 329


RESULT   7
FOMVRV
gag polyprotein - radiation murine leukemia virus
N;Contains: core protein p15; core shell protein p30; inner coat protein p12; nucleop
C;Species: radiation murine leukemia virus
C;Date: 31-Mar-1989 #sequence_revision 31-Mar-1989 #text_change 16-Jul-1999
C;Accession: A26183
R;Merregaert, J.; Janowski, M.; Reddy, E.P.
Virology 158, 88-102, 1987
A;Title: Nucleotide sequence of a radiation leukemia virus genome.
A;Reference number: A94362; MUID:87207680
A;Accession: A26183
A;Molecule type: DNA
A;Residues: 1-537 <MER>
A;Cross-references: GB:K03363; GB:M18449; NID:g332032; PIDN:AAA46518.1; PID:g332033
C;Genetics:
A;Gene: gag
C;Superfamily: mammalian retrovirus gag polyprotein I
C;Keywords: core protein; inner coat protein; nucleoprotein; polyprotein
F;1-129/Product: core protein p15 #status predicted <P15>
F;130-214/Product: inner coat protein p12 #status predicted <P12>
F;215-477/Product: core shell protein p30 #status predicted <P30>
F;478-537/Product: nucleoprotein p10 #status predicted <P10>


  Query Match             6.3%;  Score 135;  DB 1;  Length 537;
  Best Local Similarity   19.1%;  Pred. No. 0.32;
  Matches   97;  Conservative   58;  Mismatches  173;  Indels  180;  Gaps   20;

Qy      21 PGLHPHRS--------EATAGRSPP------TPTVTLGPDCP-----PPPPPPPPNNNNN 61
           || | |           || |   |        :| ::| |  |      |   |||
Db      71 PGPHGHPDQVPYIVTWEAIAYEPPSWVKPFVSPKLSLSPTAPILPSGPSTQPPP------ 124

Qy      62 NNSKHTGHKSACVPNMTERRRDELSEEINNLREKVMKQSEENNNLQSQVQKLTEENTTLR 121
                  :||  | :|           ::: :  |    ::|    :  |||:
Db     125 --------RSALYPALTP-----------SIKPRPSKPQVLSDNGGPLIDLLTEDPPPYG 165

Qy     122 EQVEPTPEDEDDDIELRGAAAAAAP-----------PPPIEEECPEDLPEKFDGNPDM- 168
           ||   :|: : |  |    :   ||             ||  :       | :  ||  :
Db     166 EQGPSSPDGDGDREEATYTSEIPAPSPMVSRLRGKRDPPAADSTTSRAFPLRLGGNGQLQ 225

Qy     169 LAPFMAQCQIFMEKSTRDFSVDRVRV-CFVTSMMTGRAARWASAKLERSHYLMHNYPAFM 227
             || :      : :   || |  ::   : |::|     |   :              :
Db     226 YWPFSSSDLYNWKNNNPSFSEDPGKLTALIESVLTTHQPTWDDCQQ------------L 272

Qy     228 MEMKHVFEDPQRREVAKRKIRRLRQGMGSVI--DYSNAFQMIAQDLDWNEP--------- 276
           :      |: ||  :  ||  |   | : :  : ::|| :   | |:  |
Db     273 LGTLLTGEEKQRVLLEARKAVRGNDGRPTQLPNEVNSAFPLERPDWDYTTPEGRNHLVLY 332
```

```
Qy    277 --------------------------------ALIDQYHEGL--------SDHIQE--- 292
                                          | :::  |           || ||
Db    333 RQLLLAGLQNAGRSPTNLAKVKGITQGPNESPSAFLERLKEAYRRYTPYDPEDHGQETSV 392

Qy    293 -------------ELSHLE--VAKSLSALIGQCIHI----------------------E 314
                       :|  ||   :|:|  |: :   |                       |
Db    393 SMSFIWQSAPDIGRKLERLEDLKSKTLRDLVREAEKIFNKRETPEEREERFRRETEENEE 452

Qy    315 RRLARAAAARKPRSPPR----ALVLPHIASHHQVDPTEPVGGARMRLTQEEKERRRKLNL 370
          || |      | |   |    : :|  : :  : |     || | |  | :|::
Db    453 RRRAEDEQREKERDRRRQREMSKLLATVVTGQRQDRQ---GGERKR-PQLDKDQ------ 502

Qy    371 CLYCGTGGHYADNCPAKASKSSPAGNSP 398
          | ||   ||:| :|| |     | |  |
Db    503 CAYCKEKGHWAKDCPKK--PRGPRGPRP 528


RESULT  8
T30349
structural protein pp78-81 - Lymantria dispar nuclear polyhedrosis virus
C;Species: Lymantria dispar nuclear polyhedrosis virus, LdMNPV
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C;Accession: T30349
R;Kuzio, J.; Pearson, M.N.; Harwood, S.H.; Funk, C.J.; Evans, J.T.; Slavicek, J.M.; Rohr
Virology 253, 17-34, 1999
A;Title: Sequence and analysis of the genome of a baculovirus pathogenic for Lymantria d
A;Reference number: Z20836; MUID:99124785
A;Accession: T30349
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-555 <KUZ>
A;Cross-references: EMBL:AF081810; PIDN:AAC70187.1


  Query Match             6.3%;  Score 134.5;  DB 2;  Length 555;
  Best Local Similarity   20.7%;  Pred. No. 0.35;
  Matches   75;  Conservative   51;  Mismatches  129;  Indels  107;  Gaps   16;

Qy     21 PGLHPHRSEATAGRSPP------TPTVTLGPDCPPPPPPPPPNNNNNNNSKHTGHKSACV 74
          |   | | |:  | | |      |||     |  |||||||||         :    ||: |
Db    214 PASEPARQESPIGSSAPEPIRQETPTGLFAPPPPPPPPPPPP------PPEPLQQKSSAV 267

Qy     75 ----------PNMTERRRDELSEEINNLREKVMKQSEENNNLQSQVQK------LTEENT 118
                    |   :    |: :|:   | |   :    : | :::::       ||
Db    268 PPPPPPPLPPPGAPDDPFGEIGQEV---RPKPAAERAPTDALFAEIRRGVQLKPATERAP 324

Qy    119 T---------LRE--QVEPTPEDED-----------------DDIELRGAAAAAAPPP- 148
          |         :|:  :::|  |  |                 | |:|:  |  |  ||
Db    325 TYTPDALFAEIRQGVKLKPAAERADEPPKPSSRAPLLLEIENRDKIKLKKVAPRATEPPA 384

Qy    149 -----PIE-------EECPEDLPEKFDGNPDMLAPFMAQCQIFMEKSTRDFSVD------ 190
               |:        |       |: | |             :   |  |   |
Db    385 SATNNPLMQLLNKRLESMKMSSAEESDAN------------YTSSSWSDAEDDSLLRDA 431

Qy    191 -RVRVCFVTSMMTGR-----AARWASAKLERSHYLMHNYPAFMMEMKHVFEDPQRREVAK 244
           |:::  :   ::       | | | |||: :   :    |  :|  :   |  :  |
Db    432 LRIKLALLGPRLSESERKRIAKRLAGAKLKSAEKTLDELQARAIEPDNPLLSPPYQLTAP 491

Qy    245 RKIRRLRQGMGSVID------YSNAFQMI--AQDLDWNEPALIDQYHEGLSDHIQEELSH 296
            :  |:    :|:|      |  | : :  |  :|   |:|  : |: :| ::: :
Db    492 LYLHDLKLFESAVLDLFNRGAYETALEKLEEALQVDLQAPSL-QRMHDDISTYVKLQKKR 550

Qy    297 LE 298
          ||
Db    551 LE 552


RESULT  9
A59310
unconventional myosin heavy chain - maize
N;Alternate names: MYO1
C;Species: Zea mays (maize)
C;Date: 09-Jun-2000 #sequence_revision 09-Jun-2000 #text_change 08-Sep-2000
C;Accession: A59310
R;Liu, L.; Pesacreta, T.C.
submitted to GenBank, May 1999
A;Reference number: A59310
A;Accession: A59310
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-1529 <LIU>
A;Cross-references: GB:AF104924; NID:g4733890; PIDN:AAD17931.2; PID:g4733891
C;Genetics:
A;Gene: MYO1
C;Superfamily: myosin MYO2; myosin motor domain homology
F;65-719/Domain: myosin motor domain homology <MMO>


  Query Match             6.3%;  Score 134;  DB 2;  Length 1529;
  Best Local Similarity   22.3%;  Pred. No. 1.2;
  Matches   73;  Conservative   55;  Mismatches  128;  Indels   72;  Gaps   14;

Qy     78 TERRRDELSEEINNLREKVMKQSEENNNLQSQVQKLTEENTTLREQ---VEP-------- 126
          :| | :|| ::  :  :|: :  |  : |: :   :  ||  ||:|   : |
Db   1005 SELRNEELIKKFESAEKKIEQLQETVHRLEEKATNMESENKVLRQQAVAISPTSKSLAAY 1064

Qy    127 --------TPEDEDDDIELRGAAAAAAPPPPI---------EEECPEDLPEKFDGNPDML 169
                  |||: :    | |   ::    ||          ||:  : | ||   | |:|
Db   1065 PKSPFQLKTPENGN---ALNGEVKSSPDITPILPNPKELEAEEKPQKSLNEKQQENQDLL 1121

Qy    170 APFMAQ-----------CQIF-MEKSTRDFSVDRVRVCFVTSMMTGRAARWASAKLERS 216
            ::|              | | |:|  | |   : |  :|  :      ::
Db   1122 IKCVSQDLGFSSGKPIAACLIYRCLLHWRSFEVERTGV-FDRIIQTIGSAIESQDNNDKL 1180

Qy    217 HYLMHNYPAFMMEMKHVFE-------DPQRREVAKRKIRRLRQGM-GSVIDYSNAF---Q 265
            | : |    :: ::   :        ||||  :     |:  ||  |      ||   :
Db   1181 AYWLSNSSTLLLLLQRTLKTTGAAGFTPQRRRSSAASFGRVFSGMRASPQSAGRAFMGSR 1240

Qy    266 MIA-----QDLDWNEPAL-----IDQYHEGLSDHIQEELSHLEVAKSLSALIGQCIHIER 315
            :|      : ::   |||     :  : | :   |:: |      | :| |:| ||
Db   1241 LIGGLGDLRQVEAKYPALLFKQQLTAFLEKIYGMIRDNLK-----KEISPLLGLCIQAP- 1294

Qy    316 RLARAAAARKPRSPPRALVLPHIASHHQ 343
          | :||:  :  ||   ||    : :| |
Db   1295 RTSRASLIKGSRSQANALAQQTLIAHWQ 1322


RESULT  10
S70394
gag polyprotein - Friend murine leukemia virus (strain FB29)
N;Contains: core protein p15; core shell protein p30; inner coat protein p12; nucleop
C;Species: Friend murine leukemia virus
A;Variety: strain FB29
C;Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 20-Jun-2000
C;Accession: S70394
R;Perryman, S.; Nishio, J.; Chesebro, B.
Nucleic Acids Res. 19, 6950, 1991
A;Title: Complete nucleotide sequence of Friend murine leukemia virus, strain FB29.
A;Reference number: S70393; MUID:92107687
A;Accession: S70394
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: genomic RNA
A;Residues: 1-538 <PER>
A;Cross-references: EMBL:Z11128; NID:g61547; PIDN:CAA77478.1; PID:g2654364
A;Experimental source: strain FB29
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1991
C;Genetics:
A;Gene: gag
C;Superfamily: mammalian retrovirus gag polyprotein I
C;Keywords: core protein; inner coat protein; nucleoprotein; polyprotein


  Query Match             6.1%;  Score 131.5;  DB 2;  Length 538;
  Best Local Similarity   18.8%;  Pred. No. 0.53;
  Matches   95;  Conservative   53;  Mismatches  185;  Indels  171;  Gaps   17;
```

```
Qy      21 PGLHPHRS--------EATAGRSP--------PTPTVTLGPDCPPPPPPPPPNNNNNNNS 64
Db      71 PGPHGHPDQVPYIVTWEAIAVDPPPWVRPFVHPKPPLSLPPSAPSLPPEPPLS------- 123

Qy      65 KHTGHKSACVPNMTERRRDELSEEINNLREKVMKQSEENNNLQSQVQKLTEENTTLREQV 124
Db     124 --TPPQSSLYPALT-------SPLNTKPRPQVLPDSG-----GPLIDLLTEDPPPYRDPG 169

Qy     125 EPTPEDEDDDIELRGAAAAAAP------------PPPIEEECPEDLPEKFDGNPD-MLAP 171
Db     170 PPSPDGNGDSGEVAPTEGAPDPSPMVSRLRGRKEPPVADSTTSQAFPLRLGGNGQYQYWP 229

Qy     172 FMAQCQIFMEKSTRDFSVDRVRV-CFVTSMMTGRAARWASAKLERSHYLMHNYPAFMMEM 230
Db     230 FSSSDLYNWKNNNPSFSEDPAKLTALIESVLLTHQPTWDDCQQ------------LLGT 276

Qy     231 KHVFEDPQRREVAKRKIRRLRQGMGSVI--DYSNAFQMIAQDLDWN-------------- 274
Db     277 LLTGEEKQRVLLEARKAVRGEDGRPTQLPNDINDAFPLERPDWDYNTQRGRNHLVHYRQL 336

Qy     275 ----------------------------EPALIDQYHEGL-------------------- 286
Db     337 LLAGLQNAGRSPTNLAKVKGITQGPNESPSAFLERLKEAYRRYTPYDPEDPGQETNVAMS 396

Qy     287 -----SDHIQEELSHLE--VAKSLSALIGQCIHI----------------------ERRL 317
Db     397 FIWQSAPDIGRKLERLEDLKSKTLGDLVREAEKIFNKRETPEEREERIRRETEEKEERRR 456

Qy     318 ARAAAARKPRSPPRALVLPHIASHHQVDPTEPVGGARMRLTQEEKERRRKL---NLCLYC 374
Db     457 AEDVQREKERDRRR---------HREMSKLLATVVSGQRQDRQGGERRRPQLDHDQCAYC 507

Qy     375 GTGGHYADNCPAKASKSSPAGNSP 398
Db     508 KEKGHWARDCPKK--PRGPRGPRP 529
```

RESULT 11
T18256
probable serine/threonine-specific protein kinase (EC 2.7.1.-) - yeast (Candida albicans
C;Species: Candida albicans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T18256
R;Leberer, E.; Harcus, D.; Broadbent, I.D.; Clark, K.L.; Dignard, D.; Ziegelbauer, K.; S
Proc. Natl. Acad. Sci. U.S.A. 93, 13217-13222, 1996
A;Title: Signal transduction through homologs of the Ste20p and Ste7p protein kinases ca
A;Reference number: Z18843; MUID:97075145
A;Accession: T18256
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1230 <LEB>
A;Cross-references: EMBL:L47210; NID:g2276410; PID:g2286042; PIDN:AAB65439.1
C;Keywords: phosphotransferase; serine/threonine-specific protein kinase

  Query Match 6.1%; Score 131.5; DB 2; Length 1230;
  Best Local Similarity 21.5%; Pred. No. 1.4;
  Matches 64; Conservative 51; Mismatches 121; Indels 61; Gaps 13;

```
Qy      17 GGQDPGLHPHR-SEATAGRSPPTPTVTLGPDCPPPPPPPPPNNNNNNNSKHTGHKSA--- 72
Db     770 GGENNALPKQRINEFKAHRAPPPP--PSAPPAPPVPPAPPANLLSEQTSEIPQQRTAPSQ 827

Qy      73 ------CVPNMTERRRDELSEEINNLREKVMKQSEENNNLQSQVQKLTEENTTLREQVEP 126
Db     828 ALADVTAPTNIYEIQQTKYQEAQQKLREKKARELEE-------IQRLREKNERQNRQQET 880

Qy     127 TPEDEDDDIELRGAAAAAAPPPPIEEECPEDLPEKFDGNPD-MLAPFMAQ---------- 175
Db     881 GQNNAD----TASGGSNIAPPVPVPNKKP---PSGSGGGRDAKQAALIAQKKREEKKRKN 933

Qy     176 CQIFMEKST-------RDFSVDRVRVCFVTSMMTGRAARWASAKLERSHYLMHNYPAFMM 228
Db     934 LQIIAKLKTICNPGDPNELYVDLVKI----------GQGASGGVFLAHDVRDKSNIVAI 982

Qy     229 EMKHVFEDPQRREVAKRKIRRLRQGMGSVIDYSNAFQMIAQDLDWNEPALIDQYHEG 285
Db     983 KQMNLEQQPKKELIINEILVMKGSSHPNIVNFIDSY-LLKGDL-W----VIMEYMEG 1033
```

RESULT 12
T18259
serine/threonine protein kinase homolog - yeast (Candida albicans)
C;Species: Candida albicans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 11-May-2000
C;Accession: T18259
R;Kohler, J.R.; Fink, G.R.
Proc. Natl. Acad. Sci. U.S.A. 93, 13223-13228, 1996
A;Title: Candida albicans strains heterozygous and homozygous for mutations in mitoge
A;Reference number: Z11118; MUID:97075146
A;Accession: T18259
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1230 <KOH>
A;Cross-references: EMBL:U73457; NID:g1657953; PID:g1737181; PIDN:AAB38875.1
C;Genetics:
A;Note: CST20

  Query Match 6.1%; Score 131.5; DB 2; Length 1230;
  Best Local Similarity 21.5%; Pred. No. 1.4;
  Matches 64; Conservative 51; Mismatches 121; Indels 61; Gaps 13;

```
Qy      17 GGQDPGLHPHR-SEATAGRSPPTPTVTLGPDCPPPPPPPPPNNNNNNNSKHTGHKSA--- 72
Db     770 GGENNALPKQRINEFKAHRAPPPP--PSAPPAPPVPPAPPANLLSEQTSEIPQQRTAPSQ 827

Qy      73 ------CVPNMTERRRDELSEEINNLREKVMKQSEENNNLQSQVQKLTEENTTLREQVEP 126
Db     828 ALADVTAPTNIYEIQQTKYQEAQQKLREKKARELEE-------IQRLREKNERQNRQQET 880

Qy     127 TPEDEDDDIELRGAAAAAAPPPPIEEECPEDLPEKFDGNPD-MLAPFMAQ---------- 175
Db     881 GQNNAD----TASGGSNIAPPVPVPNKKP---PSGSGGGRDAKQAALIAQKKREEKKRKN 933

Qy     176 CQIFMEKST-------RDFSVDRVRVCFVTSMMTGRAARWASAKLERSHYLMHNYPAFMM 228
Db     934 LQIIAKLKTICNPGDPNELYVDLVKI----------GQGASGGVFLAHDVRDKSNIVAI 982

Qy     229 EMKHVFEDPQRREVAKRKIRRLRQGMGSVIDYSNAFQMIAQDLDWNEPALIDQYHEG 285
Db     983 KQMNLEQQPKKELIINEILVMKGSSHPNIVNFIDSY-LLKGDL-W----VIMEYMEG 1033
```

RESULT 13
T01815
hypothetical protein T27D20.19 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 26-Feb-1999 #sequence_revision 26-Feb-1999 #text_change 24-Mar-1999
C;Accession: T01815
R;Edwards, J.; Wollam, C.; Dubbelde, C.
submitted to the EMBL Data Library, August 1998
A;Description: The sequence of A. thaliana T27D20.
A;Reference number: Z14441
A;Accession: T01815
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-334 <EDW>
A;Cross-references: EMBL:AF076274; NID:g3293583; PID:g3377854
A;Experimental source: cultivar Columbia
C;Genetics:
A;Map position: 4
A;Introns: 79/1; 107/2; 133/2
A;Note: T27D20.19

```
  Query Match             6.1%;  Score 130;  DB 2;  Length 334;
  Best Local Similarity   22.4%;  Pred. No. 0.38;
  Matches   76;  Conservative   56;  Mismatches  109;  Indels   98;  Gaps   20;

Qy      85 LSEEINNLREKV--MKQSEENNNL-QSQVQKLTEENTTLREQVEPTPE-DEDDDIELRGA 140
           |:||:  ||:::  ::|      | | | ::        ||: |  |: |::||
Db       4 LTEELRGLRKRLGRLEQPYPRVELGQDQARR------NAREEAEEFPDLDDEDD------ 51

Qy     141 AAAAAPPPPI--EEECPED----------LPEKFDGNPDMLAPFMAQCQ-IF----MEKS 183
             || |:  : :  ||          | |   ||::   :  : : ||    ||
Db      52 ----PPPDPLRRQHQRREDPLMRNRARGVEPRK--SNPEVYMDWERRMEYIFECYGMENL 105

Qy     184 TRDFSVDRVRVCFVTSMMTGRAARWASAKLERSHYLMHNYPAFMMEMKHVFEDPQRREVA 243
           |  : || |  |      |    |   |     ||:
Db     106 GRTVA-DRRRKRF------GPINTWGDMK-----YLL---------------------- 130

Qy     244 KRKIRRLRQGMGSVIDYSNAFQ--MIAQDLDWNEPALIDQYHEGLSDHIQEELSHLEVAK 301
             ::| | ||  :| :|   |:  | | :|: :: ||: |: :|| : || ::   :
Db     131 --RLRMLSQGTRTVDEYFEEFEKLMNALELEESDEALMAQFIDGLQERIQRKVERAQY-N 187

Qy     302 SLSALIGQCIHIERRLARAAA----------ARKPRSPPRALVLPHIASHHQV--DPT 347
           |  |:   : :|::: | |:            |   |:  :: |   : | |:
Db     188 GLHELLHLAVQVEQQIRRKASLSNRSRNNTPWNASNNRAMDKSKV---VESDHRFKNKSN 244

Qy     348 EPVGGARMRLTQ--EEKERRRKLNLCLYCGTGGHYADNC 384
           |    :| :| :     : | :   |  |    || |  |
Db     245 EAPKTSRPKLGKFPSTNQSRSRYITCFKCQGRGHMAREC 283


RESULT  14
FOMVMB
gag polyprotein - murine leukemia virus (strain BM5 ECO)
N;Contains: core protein p15; core shell protein p30; inner coat protein p12; nucleoprot
C;Species: murine leukemia virus
C;Date: 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change 16-Jul-1999
C;Accession: A40416
R;Chattopadhyay, S.K.; Sengupta, D.N.; Fredrickson, T.N.; Morse III, H.C.; Hartley, J.W.
J. Virol. 65, 4232-4241, 1991
A;Title: Characteristics and contributions of defective, ecotropic, and mink cell focus-
A;Reference number: A40416; MUID:91303677
A;Accession: A40416
A;Molecule type: DNA
A;Residues: 1-537 <CHA>
A;Cross-references: GB:M64095; NID:g332014; PIDN:AAA46510.1; PID:g332015
C;Genetics:
A;Gene: gag
C;Superfamily: mammalian retrovirus gag polyprotein I
C;Keywords: core protein; inner coat protein; nucleoprotein; polyprotein
F;1-129/Product: core protein p15 #status predicted <CP1>
F;130-214/Product: inner coat protein p12 #status predicted <ICP>
F;215-477/Product: core shell protein p30 #status predicted <CSP>
F;478-537/Product: nucleoprotein p10 #status predicted <NP1>


  Query Match             6.1%;  Score 130;  DB 1;  Length 537;
  Best Local Similarity   18.5%;  Pred. No. 0.65;
  Matches   93;  Conservative   53;  Mismatches  186;  Indels  172;  Gaps   17;

Qy      21 PGLHPHRS--------EATAGRSPP------TPTVTLGPDCP-----PPPPPPPPNNNNN 61
           || | |          || |   ||      :| ::| |  |     |   |||
Db      71 PGPHGHPDQVPYIVTWEAIAYEPPPWVKPFVSPKLSLSPTAPILPSGPSTQPPP------ 124

Qy      62 NNSKHTGHKSACVPNMTERRRDELSEEINNLREKVMKQSEENNNLQSQVQKLTEENTTLR 121
                 :||  | |           |      |   |::   |   :  |||:
Db     125 --------RSALYPAFT--------PSIKPRPSKPQVLSDDGGPL---IDLLTEDPPPYG 165

Qy     122 EQVEPTPEDEDDDIELRGAAAAAAP------------PPPIEEECPEDLPEKFDGNPDM- 168
           ||   :|: : |  |    :   ||            ||  :       | :  ||  :
Db     166 EQGPSSPDGDGDREEATSTSEIPAPSPMVSRLRGKRDPPAADSTTSRAFPLRLGGNGQLQ 225

Qy     169 LAPFMAQCQIFMEKSTRDFSVDRVRV-CFVTSMMTGRAARWASAKLERSHYLMHNYPAFM 227
             || :      : :   || |  ::   : |::|     |   :              :
Db     226 YWPFSSSDLYNWKNNNPSFSEDPGKLTALIESVLTTHQPTWDDCQQ------------L 272

Qy     228 MEMKHVFEDPQRREVAKRKIRRLRQGMGSVI--DYSNAFQMIAQDLDWNEP--------- 276
           :      |: ||  :  ||  |   |  : :  : ::|| :   | |:  |
Db     273 LGTLLTGEEKQRVLLEARKAVRGNDGRPTQLPNEVNSAFPLERPDWDYTTPEGRNHLVLY 332

Qy     277 --------------------------------ALIDQYHEGLSDH-------------- 289
                                            | :::  |    :
Db     333 RQLLLAGLQNAGRSPTNLAKVKGITQGPNESPSAFLERLKEAYRRYTPYDPEDPGQETNV 392

Qy     290 -----------IQEELSHLE--VAKSLSALIGQCIHI----------------------E 314
                       |  :|  ||   :|:|  |: :   |                      |
Db     393 SMSFIWQSAPAIGRKLERLEDLKSKTLGDLVREAEKIFNKRETPEEREERIRRETEEKEE 452

Qy     315 RRLARAAAAARKPRSPPRALVLPHIASHHQVDPTEPVGGARMRLTQEEKERRRKLNLCLYC 374
           || |      | |   |   :  : :       :   |   |  | :|::      | ||
Db     453 RRRAGDEQREKERDRRRQREMSKLLATVVTGQRQDRQGGERRRPQLDKDQ------CAYC 506

Qy     375 GTGGHYADNCPAKASKSSPAGNSP 398
             ||:| :|| |     | |  |
Db     507 KEKGHWAKDCPKK--PRGPRGPRP 528


RESULT  15
T19673
hypothetical protein C33B4.3 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T19673
R;Coles, L.
submitted to the EMBL Data Library, February 1995
A;Reference number: Z19160
A;Accession: T19673
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1110 <WIL>
A;Cross-references: EMBL:Z48367; PIDN:CAA88324.1; GSPDB:GN00020; CESP:C33B4.3
A;Experimental source: clone C33B4
C;Genetics:
A;Gene: CESP:C33B4.3
A;Map position: 2
A;Introns: 20/3; 110/3; 144/3; 341/1; 455/1; 546/1; 653/3; 934/2; 983/3; 1035/1


  Query Match             6.0%;  Score 129;  DB 2;  Length 1110;
  Best Local Similarity   21.7%;  Pred. No. 1.8;
  Matches   34;  Conservative   21;  Mismatches   42;  Indels   60;  Gaps    3;

Qy      36 PPTPTVTLGPDCPPPPPPPPP---------------------------------------- 56
           || |   :    |||||||||
Db     796 PPPPLPPISSGAPPPPPPPPPGGLMHVAASAPVLMSNSKGISADALKSVQLKKAEPRETS 855

Qy      57 ----NNNNNNNSKHTGHKSACVPNMTERRRDELSEEIN----------------NLREKV 96
               :||||||:  |      : |   :|| ::: :::                :|| |
Db     856 AASVSNNNNNNNNSTTDFQMDLKNALAKRRSKVAHDVDEDEERESRFEGLSLRETVRENV 915

Qy      97 MKQSEENNNLQSQVQKLTEENTTLREQVEPTPEDEDD 133
           ::: :   |:   | |     |: |  :||:  :| |
Db     916 VERGKGIQNI-GIVNKKDSGYTSSRTSLEPSESEEKD 951


Search completed: October  4, 2002, 15:39:11
Job time: 53 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: October 4, 2002, 15:38:18 ; Search time 31.76 Seconds
(without alignments)
1402.412 Million cell updates/sec

Title: US-09-631-863A-2
Perfect score: 2139
Sequence: 1 MRNKRVLKTKKRRSGRGGQD.....DNCPAKASKSSPAGNSPAPL 401

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_032802:*
1: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1802	84.2	342	21	AAB42148	Human ORFX ORF1912
2	245.5	11.5	1607	22	ABG15099	Novel human diagno
3	191	8.9	502	22	ABG09602	Novel human diagno
4	159	7.4	146	20	AAY26051	Human alpha helica
5	159	7.4	146	22	AAB60492	Human cell cycle a
6	156	7.3	520	22	AAU15111	Protein encoded by
7	155	7.2	126	20	AAY25512	Human mature alpha
8	149.5	7.0	110	20	AAY26054	Human Zalpha1 epit
9	149.5	7.0	168	21	AAY94674	Human zsig83 matur
10	149.5	7.0	184	21	AAY94673	Human zsig83 prote
11	139.5	6.5	142	22	ABG12205	Novel human diagno
12	139	6.5	113	22	AAB60475	Human cell cycle a
13	137	6.4	144	20	AAY59927	Human myometrium t
14	135.5	6.3	1230	19	AAW48895	Candida albicans C
15	132	6.2	557	19	AAY20666	Human neurofilamen
16	130	6.1	676	22	ABB59726	Drosophila melanog
17	129	6.0	872	22	AAB90753	Human shear stress
18	126.5	5.9	567	22	ABB64078	Drosophila melanog
19	126.5	5.9	567	22	ABB65901	Drosophila melanog
20	126.5	5.9	567	22	ABB66007	Drosophila melanog
21	126	5.9	1072	22	AAG70871	C albicans apoptos
22	125	5.8	905	18	AAW31186	Human p160 polypep
23	125	5.8	1135	18	AAW31185	Human p160 polypep
24	123	5.8	874	22	ABB62601	Drosophila melanog
25	123	5.8	1321	22	ABG25416	Novel human diagno
26	122.5	5.7	1316	22	ABG22997	Novel human diagno
27	122.5	5.7	2348	22	ABG10929	Novel human diagno
28	121.5	5.7	486	22	AAU30801	Novel human secret
29	121.5	5.7	529	22	ABB12368	Human bone marrow
30	120.5	5.6	1157	22	ABG19976	Novel human diagno
31	120.5	5.6	1743	22	ABG10928	Novel human diagno
32	120	5.6	80	20	AAY26056	Human Zalpha1 epit
33	120	5.6	954	22	AAU14615	Novel bone marrow
34	119.5	5.6	622	22	ABB62816	Drosophila melanog
35	119.5	5.6	2954	20	AAY01632	Amino acid sequenc
36	119	5.6	505	19	AAW46889	Human Neural-Wisko
37	119	5.6	505	22	AAM52316	Human N-WASP prote
38	119	5.6	505	22	AAG67338	Amino acid sequenc
39	119	5.6	582	22	ABG16505	Novel human diagno
40	118.5	5.5	2293	21	AAY69197	Amino acid sequenc
41	118	5.5	818	18	AAW13386	Human protein ubiq
42	118	5.5	2175	22	ABB65698	Drosophila melanog
43	117.5	5.5	501	19	AAW46890	Rat Neural-Wiskott
44	117.5	5.5	501	22	AAM52319	Rat N-WASP protein
45	117.5	5.5	501	22	AAG67355	Amino acid sequenc

ALIGNMENTS

RESULT 1
AAB42148
ID AAB42148 standard; Protein; 342 AA.
XX
AC AAB42148;
XX
DT 08-FEB-2001 (first entry)
XX
DE Human ORFX ORF1912 polypeptide sequence SEQ ID NO:3824.
XX
KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW vulnerary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW thrombosis; contraceptive.
XX
OS Homo sapiens.
XX
PN WO200058473-A2.
XX
PD 05-OCT-2000.
XX
PF 31-MAR-2000; 2000WO-US08621.
XX
PR 31-MAR-1999; 99US-0127607.

```
PR   02-APR-1999;   99US-0127636.
PR   05-APR-1999;   99US-0127728.
PR   30-MAR-2000; 2000US-0540763.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Shimkets RA,  Leach M;
XX
DR   WPI; 2000-602362/57.
DR   N-PSDB; AAC76357.
XX
PT   Novel nucleic acids and peptides derived from open reading frame X,
PT   useful for treating e.g. cancers, proliferative disorders,
PT   neurodegenerative disorders and cardiovascular disease -
XX
PS   Claim 11; Page 2977-2978; 5507pp; English.
XX
CC   AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC   which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC   sequences have activities such as: cytostatic; hepatotropic; vulnerary;
CC   antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
CC   osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
CC   immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
CC   antidiabetic; hypotensive; dermatological; immunosuppressive;
CC   antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
CC   antithyroid; and antianaemic. The sequences can be used for determining
CC   the presence of or predisposition to, or preventing or treating
CC   pathological conditions associated with an ORFX-associated disorder. The
CC   nucleic acids can be used to express ORFX proteins in gene therapy
CC   vectors. The proteins and nucleic acids may be used to treat cancers,
CC   proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC   graft vs host disease, cardiovascular disease, diabetes mellitus,
CC   hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC   erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC   bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC   allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC   nocturnal haemoglobinuria, antiinflammatory disease; to enhance
CC   coagulation; to inhibit thrombosis; and as a contraceptive.
XX
SQ   Sequence   342 AA;
```

```
  Query Match             84.2%;  Score 1802;  DB 21;  Length 342;
  Best Local Similarity   100.0%;  Pred. No. 1.8e-137;
  Matches  342;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      60 NNNNSKHTGHKSACVPNMTERRRDELSEEINNLREKVMKQSEENNNLQSQVQKLTEENTT 119
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 nnnnskhtghksacvpnmterrrdelseeinnlrekvmkqseennnlqsqvqklteentt 60

Qy     120 LREQVEPTPEDEDDDIELRGAAAAAAPPPPIEEECPEDLPEKFDGNPDMLAPFMAQCQIF 179
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 lreqveptpededddielrgaaaaaapppppieeecpedlpekfdgnpdmlapfmaqcqif 120

Qy     180 MEKSTRDFSVDRVRVCFVTSMMTGRAARWASAKLERSHYLMHNYPAFMMEMKHVFEDPQR 239
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 mekstrdfsvdrvrvcfvtsmmtgraarwasaklershylmhnypafmmemkhvfedpqr 180

Qy     240 REVAKRKIRRLRQGMGSVIDYSNAFQMIAQDLDWNEPALIDQYHEGLSDHIQEELSHLEV 299
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 revakrkirrlrqgmgsvidysnafqmiaqdldwnepalidqyheglsdhiqeelshlev 240

Qy     300 AKSLSALIGQCIHIERRLARAAAARKPRSPPRALVLPHIASHHQVDPTEPVGGARMRLTQ 359
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 akslsaligqcihierrlaraaaarkprsppralvlphiashhqvdptepvggarmrltq 300

Qy     360 EEKERRRKLNLCLYCGTGGHYADNCPAKASKSSPAGNSPAPL 401
           |||||||||||||||||||||||||||||||||||||||||||
Db     301 eekerrrklnlclycgtgghyadncpakasksspagnspapl 342
```

```
RESULT   2
ABG15099
```

```
ID   ABG15099 standard; Protein; 1607 AA.
XX
AC   ABG15099;
XX
DT   18-FEB-2002  (first entry)
XX
DE   Novel human diagnostic protein #15090.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   N-PSDB; AAS79286.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
XX
PS   Claim 20; SEQ ID No 45458; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human
CC   diagnostic amino acid sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   1607 AA;
```

```
  Query Match             11.5%;  Score 245.5;  DB 22;  Length 1607;
  Best Local Similarity   25.1%;  Pred. No. 7.9e-11;
  Matches   99;  Conservative   54;  Mismatches  149;  Indels   93;  Gaps   15;

Qy       9 TKKRRSG---RGGQDPGLHPHRSEATAGRSPPT---------PTVTLGPDCPPPPPPP-- 54
           | :  ||   ||   |   |    |   : ||:          ||  |  |   |   |
Db     282 tteatsgsgvrgeagpasgp----aqekkeppsgplqemeelptdll-qdmeepssgprk 336

Qy      55 ----PPNN-----NNNNNSKHTGHKSACVPNMTERRRDELSEEINNLREK------VMKQ 99
               |||:       : |  |            : | | ||   : ::|         ::
Db     337 eiedppndllqdleescngsh-----------qargdplsgasdrmkeasvnpsgaree 384
```

```
Qy    100 SEENNNLQS-------QVQKLTEENTT-----------------LREQVEPTPEDEDDD 134
Db    385 qeahtdlkesgreetpqeqnqtehstaelmamvrsiislyfrmqdlkeq-----qrvaee 439

Qy    135 IELRGAAAAAAPPPPIEEECPEDLPEKFDGNPDMLAPFMAQCQIFMEKSTRDFSVDRVRV 194
Db    440 ilikginagqlp-----------apkhfsgdrrefhefivlcqltlqsyprmfyndrlrv 488

Qy    195 CFVTSMMTGRAARWASAKLERSHYLMHNYPAFMMEMKHVFEDPQRREVAKRKIRRLRQGM 254
Db    489 gyvinhlsglalewakallqenspligdfpafleamsevfeyrqalrvaeeamftirqgg 548

Qy    255 GSVIDYSNAFQMIAQDLDWNEPALIDQYHEGLSDHIQEELSHLEVAKSLSALIGQCIH-I 313
Db    549 rsateyidefqslvpilgwpdevlqahlcqglneeirhyl--frvpqpdsrrqsdcahpa 606

Qy    314 ERRLA---RAAAARKPRSPPRALVLPHIASHHQVD 345
Db    607 nrreagreksyaqaaprgpppepdldrltcsrevd 641


RESULT   3
ABG09602
ID   ABG09602 standard; Protein; 502 AA.
XX
AC   ABG09602;
XX
DT   13-FEB-2002  (first entry)
XX
DE   Novel human diagnostic protein #9593.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   N-PSDB; AAS73789.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
XX
PS   Claim 20; SEQ ID No 39961; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human
CC   diagnostic amino acid sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   502 AA;


  Query Match             8.9%;  Score 191;  DB 22;  Length 502;
  Best Local Similarity   25.2%;  Pred. No. 4.4e-07;
  Matches   55;  Conservative   43;  Mismatches  110;  Indels   10;  Gaps    3;

Qy    158 LPEKFDGNPDMLAPFMAQCQIFMEKSTRDFSVDRVRVCFVTSMMTGRAARWASAKLERSH 217
Db     21 lplrragsspglmpveeemeideekqmkgflddsermaflvslhlgaaerwfilqmevge 80

Qy    218 YLMHNYPAFMMEMKHVFEDPQRREVAKRKIRRLRQGMGSVIDYSNAFQMIAQDLDWNEPA 277
Db     81 plshenksflrrsqgiydslseidilsavlchpkqgqksvrqyatdflllarhlswsdai 140

Qy    278 LIDQYHEGLSDHIQEELSH--LEVAKSLSALIGQCIHIERRLARAAAARKPRSPPRALVL 335
Db    141 lrtrfleglseavttkmgriflkvagslkelidrslytecqlae----ekdspgnssqvl 196

Qy    336 PHIASHHQVDPTEPVGGARMRLTQEEKERRRKLNLCLY 373
Db    197 ptackrn----neeamgnelssqqqteehqhvskrcyy 230


RESULT   4
AAY26051
ID   AAY26051 standard; Protein; 146 AA.
XX
AC   AAY26051;
XX
DT   28-SEP-1999  (first entry)
XX
DE   Human alpha helical precursor protein-1, Zalpha1.
XX
KW   Human alpha helical precursor protein-1; Zalpha1; helical cytokine;
KW   growth hormone; erythropoietin; leptin; chromosome Xq27.3; FMR1;
KW   Fragile-X syndrome; interleukin-10;  connective tissue dysfunction;
KW   abnormal proliferative disorder; cancer; epidermal system; skin tone;
KW   hypothalamic pituitary gonadal axis dysfunction; cardiovascular system;
KW   cosmetic improvement; elasticity.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Peptide          1..20
FT                    /label= signal_peptide
FT   Protein          21..146
FT                    /label= Zalpha1
FT   Region           23..37
FT                    /label= Helix_A
FT   Region           53..67
FT                    /label= Helix_B
FT   Region           82..96
FT                    /label= Helix_C
FT   Region           118..132
FT                    /label= Helix_D
XX
PN   WO9929720-A2.
XX
PD   17-JUN-1999.
XX
PF   10-DEC-1998;   98WO-US26273.
XX
PR   10-DEC-1997;   97US-0987926.
XX
```

```
PA   (ZYMO ) ZYMOGENETICS INC.
XX
PI   Conklin DC,  Lok S,  Parrish J;
XX
DR   WPI; 1999-385572/32.
DR   N-PSDB; AAX80687.
XX
PT   Mammalian alpha helical protein-1, designated Zalpha1
XX
PS   Example 2; Pages 60-61; 73pp; English.
XX
CC   The present sequence represents an alpha helical precursor protein-1
CC   designated Zalpha1 from human pituitary gland cDNA library. Zalpha1
CC   is predicted to be a four-helical protein similar to the family of
CC   helical cytokines represented by growth hormone, erythropoietin, leptin
CC   and interleukin-10. Zalpha1 gene was mapped to chromosome Xq27.3, in
CC   close proximity to FMR1, a gene linked to Fragile-X syndrome. Its
CC   transcription levels were found to be reduced or absent in Fragile-X
CC   patients. Zalpha1 transcripts were found at high levels in pituitary and
CC   aorta, and lower levels in brain, kidney, pancreas, prostrate, etc.
CC   It is supposed to be responsible for the hypothalamic-pituitary-
CC   gonadal axis dysfunction and connective tissue dysfunctions in Fragile-X
CC   patients. Zalpha1 protein may be useful in the treatment of Fragile-X
CC   syndrome and abnormal proliferative disorders e.g. cancer. It can also be
CC   used for the growth, differentiation, maintenance and survival of
CC   connective tissues, particularly cardiovascular and epidermal systems
CC   and in imparting cosmetic improvements to normal connective tissues such
CC   as enhancement of skin tone and elasticity.
XX
SQ   Sequence   146 AA;


  Query Match             7.4%;  Score 159;  DB 20;  Length 146;
  Best Local Similarity   27.4%;  Pred. No. 3.4e-05;
  Matches   45;  Conservative   21;  Mismatches   68;  Indels   30;  Gaps    2;

Qy      83 DELSEEINNLREKVMKQSEENNNLQSQVQKLTEENTTLREQVEPTPEDEDDDIELRGAAA 142
           |||   :: |  :    | ||: |  |:: |  |  :|  || |
Db       3 delvlllhallmrhralsiensqlmeqlrllvcerasllrqvrp---------------- 46

Qy     143 AAAPPPPIEEECPEDLPEKFDGNPDMLAPFMAQCQIFMEKSTRDFSVDRVRVCFVTSMMT 202
                 ||   || |:|    |  |: |   :|  :   |  | ::| |: |::|
Db      47 ---------pscpvpfpetfngessrlpefivqtasymlvnenrfcndamkvaflisllt 97

Qy     203 GRAARWASAKLERSHYLMHNYPAFMMEMKHVF-----EDPQRRE 241
           | |  |    :|    :: :| ||: |||  |     ||    |
Db      98 geaeewvvpyiemdspilgdyrafldemkqcfgwddddeddddee 141


RESULT   5
AAB60492
ID   AAB60492 standard; Protein; 146 AA.
XX
AC   AAB60492;
XX
DT   24-APR-2001  (first entry)
XX
DE   Human cell cycle and proliferation protein CCYPR-40, SEQ ID NO:40.
XX
KW   Cell cycle and proliferation protein; CCYPR; human; agonist;
KW   antagonist; gene therapy; detection; gene therapy;
KW   transgenic animal disease model; immune disorder;
KW   developmental disorder; cell signalling disorder;
KW   cell proliferative disorder; cancer; tumour; anaemia; epilepsy;
KW   arteriosclerosis; asthma; allergy; diabetes mellitus;
KW   menstrual cycle disorder; bacterial infection.
XX
OS   Homo sapiens.
XX
PN   WO200107471-A2.
XX
PD   01-FEB-2001.
XX
PF   21-JUL-2000; 2000WO-US19948.
XX
PR   21-JUL-1999;   99US-0145075.
PR   08-SEP-1999;   99US-0153129.
PR   10-NOV-1999;   99US-0164647.
XX
PA   (INCY-) INCYTE GENOMICS INC.
XX
PI   Hillman JL,  Lal P,  Tang YT,  Yue H,  Au-Young J,  Bandman O;
PI   Azimzai Y,  Yang J,  Lu DAM,  Baughn MR,  Patterson C,  Shah P;
XX
DR   WPI; 2001-112727/12.
DR   N-PSDB; AAF59629.
XX
PT   Human cell cycle and proliferation proteins and polynucleotides are
PT   used to treat, diagnose and prevent immune, developmental and cell
PT   signaling disorders and cell proliferative disorders including cancer -
XX
PS   Disclosure; Page 150; 205pp; English.
XX
CC   Sequences AAB60453-AAB60506 represent 54 human cell cycle and
CC   proliferation proteins (CCYPR), which are encoded by AAF59590-AAF59643.
CC   CCYPR and agonists of CCYPR are used to treat diseases or conditions
CC   associated with decreased expression of functional CCYPR, while CCYPR
CC   antagonists are used to treat diseases or conditions associated with
CC   overexpression of functional CCYPR. Monoclonal or polyclonal antibodies
CC   to CCYPR may be used in enzyme-linked immunosorbent assays (ELISA) or
CC   radioimmunoasays to detect CCYPR. CCYPR itself may be used to detect
CC   compounds e.g., antibodies, oligonucleotides and proteins (receptors)
CC   that specifically bind to CCYPR, and in drug screening methods to
CC   identify compounds that modulate the activity of CCYPR. CCYPR
CC   nucleotides can be used to generate transgenic animal models of human
CC   disease, and can be used in gene therapy in target cells with genetic
CC   abnormalities with respect to the expression of CCYPR for the
CC   treatment or prevention of a disorder associated with CCYPR.
CC   Diseases which can be diagnosed, treated and prevented using CCYPR
CC   proteins, nucleic acids, agonists or antagonists include immune,
CC   developmental and cell signalling disorders, and cell proliferative
CC   disorders including cancer. Specific examples of these disorders
CC   include anaemia, epilepsy, arteriosclerosis, asthma, cancer, allergies,
CC   diabetes mellitus, disorders of the menstrual cycle and infections
CC   caused by bacteria.
XX
SQ   Sequence   146 AA;


  Query Match             7.4%;  Score 159;  DB 22;  Length 146;
  Best Local Similarity   27.4%;  Pred. No. 3.4e-05;
  Matches   45;  Conservative   21;  Mismatches   68;  Indels   30;  Gaps    2;

Qy      83 DELSEEINNLREKVMKQSEENNNLQSQVQKLTEENTTLREQVEPTPEDEDDDIELRGAAA 142
           |||   :: |  :    | ||: |  |:: |  |  :|  || |
Db       3 delvlllhallmrhralsiensqlmeqlrllvcerasllrqvrp---------------- 46

Qy     143 AAAPPPPIEEECPEDLPEKFDGNPDMLAPFMAQCQIFMEKSTRDFSVDRVRVCFVTSMMT 202
                 ||   || |:|    |  |: |   :|  :   |  | ::| |: |::|
Db      47 ---------pscpvpfpetfngessrlpefivqtasymlvnenrfcndamkvaflisllt 97

Qy     203 GRAARWASAKLERSHYLMHNYPAFMMEMKHVF-----EDPQRRE 241
           | |  |    :|    :: :| ||: |||  |     ||    |
Db      98 geaeewvvpyiemdspilgdyrafldemkqcfgwddddeddddee 141


RESULT   6
AAU15111
ID   AAU15111 standard; Protein; 520 AA.
XX
AC   AAU15111;
XX
DT   04-DEC-2001  (first entry)
XX
DE   Protein encoded by C. albicans essential gene CaYPL228W (CET1).
XX
```

KW Gene identification; essential gene; GRACE; pathogenic fungus;
KW gene replacement and conditional expression; fungal infection.
XX
OS Candida albicans.
XX
PN WO200160975-A2.
XX
PD 23-AUG-2001.
XX
PF 20-FEB-2001; 2001WO-US05551.
XX
PR 18-FEB-2000; 2000US-0183534.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Roemer T, Jiang B, Boone C, Bussey H;
XX
DR WPI; 2001-489080/53.
DR N-PSDB; AAS23439.
XX
PT Identifying genes essential to fungal metabolisms and identifying
PT potential therapeutic agents that target these genes -
XX
PS Claim 43; Page 247-248; 324pp; English.
XX
CC The present invention relates to novel methods for constructing fungal
CC strains useful for identification and validation of gene products as
CC targets for therapeutic agents, for creating a collection of identified
CC essential genes, and screening assays for the discovery of new drugs.
CC The invention provides the GRACE (gene replacement and conditional
CC expression) method for the construction of mutant organisms referred to
CC as GRACE strains of the organism. The invention can be applied to any
CC organism, particularly a pathogenic fungus e.g. Candida albicans,
CC Aspergillus fumigatus and Cryptococcus neoformans. The methods are
CC useful to identify agents that may be used in the treatment of fungal
CC infections. AAU15053-AAU15113 represent proteins encoded by C. albicans
CC essential genes.
XX
SQ Sequence 520 AA;

Query Match 7.3%; Score 156; DB 22; Length 520;
Best Local Similarity 21.6%; Pred. No. 0.00031;
Matches 78; Conservative 59; Mismatches 118; Indels 106; Gaps 18;

```
Qy       9 TKKRRSGRGGQDPGLHPHRS-----EATAGRSPPT--PTVTLGPDCPPP----------- 50
Db      24 tkiiksptayhkpsvherhsitsmlndtpsdstptkkpeptispefrkpsissltspsva 83

Qy      51 --PPPPPPNNNNNNNSKHTGHKSACVPNMTERR-------------RDELSEEINNLRE 94
Db      84 hkppplppssssvgssehssarss--paitkrnsianiidayeepatktekkaelnspk- 140

Qy      95 KVMKQS------EENNNLQSQVQKLTEENTTLREQVEPTP--EDEDDDI-ELRGAAAAAA 145
Db     141 --inqstpvpkleehendtnkvekvvdsapepkpkkepqpvfddqdddltkikklkqskk 198

Qy     146 P----PPPI-----------EEECPEDLPEKFDGNPDMLAPFMAQCQIFMEKSTRDFSVD 190
Db     199 prryetppiwaqrwvppnrqkeetnvd-----dgneaitr--lsekpvfdytttrsvdle 251

Qy     191 RVRVCFVTSM-----MTGRAARWASAKLERSHYLMHNYPAFMMEMKHVFEDPQRREVAKR 245
Db     252 ----csitgmippssitrkiaewvyanfsn-----------------veekskrnv--- 286

Qy     246 KIRRLRQGMGSVIDYSNAFQMIAQDLDWNEPALIDQYHEGLSDHIQEELSHLEVAKSLSA 305
Db     287 ---elelkfgkiidkrsgnri---dlnvvteciftdhssvffdmqveevawkeitkflde 340

Qy     306 L 306
Db     341 l 341
```

RESULT 7
AAY25512
ID AAY25512 standard; Protein; 126 AA.
XX
AC AAY25512;
XX
DT 28-SEP-1999 (first entry)
XX
DE Human mature alpha helical protein-1 Zalpha1.
XX
KW Human alpha helical protein-1; Zalpha1; helical cytokine; skin tone;
KW growth hormone; erythropoietin; leptin; interleukin-10; gene therapy;
KW chromosome Xq27.3; FMR1; Fragile-X syndrome; cosmetic improvement;
KW hypothalamic pituitary gonadal axis dysfunction; cardiovascular system;
KW abnormal proliferative disorder; cancer; connective tissue dysfunction;
KW epidermal system; elasticity.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Protein 1..127
FT /label= Zalpha1
FT /note= "mature protein fragment, no start codon given"
XX
PN WO9929720-A2.
XX
PD 17-JUN-1999.
XX
PF 10-DEC-1998; 98WO-US26273.
XX
PR 10-DEC-1997; 97US-0987926.
XX
PA (ZYMO) ZYMOGENETICS INC.
XX
PI Conklin DC, Lok S, Parrish J;
XX
DR WPI; 1999-385572/32.
DR N-PSDB; AAX80687.
XX
PT Mammalian alpha helical protein-1, designated Zalpha1
XX
PS Claim 5; Page 69; 73pp; English.
XX
CC The present sequence represents the mature protein fragment of a novel
CC alpha helical protein-1 designated Zalpha1 from human pituitary gland
CC cDNA library. Zalpha1 is predicted to be a four-helical protein similar
CC to the family of helical cytokines represented by growth hormone,
CC erythropoietin, leptin and interleukin-10. Zalpha1 gene was mapped to
CC chromosome Xq27.3, in close proximity to FMR1, a gene linked to Fragile-X
CC syndrome. Its transcription levels were found to be reduced or absent in
CC Fragile-X patients. Zalpha1 transcripts were found at high levels in
CC pituitary and aorta, and lower levels in brain, kidney, pancreas,
CC prostrate, etc. Zalpha1 is supposed to be responsible for the
CC hypothalamic-pituitary-gonadal axis dysfunction and connective tissue
CC dysfunctions in Fragile-X patients. The Zalpha1 DNA is used in gene
CC therapy for treating patients having a mutated Zalpha1 gene or lacking
CC the gene. Probes derived from Zalpha1 gene can be used to check
CC abnormalities on X chromosome. Zalpha1 protein may be useful in the
CC treatment of Fragile-X syndrome and abnormal proliferative disorders e.g.
CC cancer. It can also be used for the growth, differentiation, maintenance
CC and survival of connective tissues, particularly cardiovascular and
CC epidermal systems and in imparting cosmetic improvements to normal
CC connective tissues such as enhancement of skin tone and elasticity.
XX
SQ Sequence 126 AA;

Query Match 7.2%; Score 155; DB 20; Length 126;
Best Local Similarity 27.6%; Pred. No. 6e-05;
Matches 40; Conservative 18; Mismatches 57; Indels 30; Gaps 2;

```
Qy     102 ENNNLQSQVQKLTEENTTLREQVEPTPEDEDDDIELRGAAAAAAPPPPIEEECPEDLPEK 161
```

```
          ||: |  |:: |  |  :|  || |                              ||   ||
Db      2 ensqlmeqlrllvcerasllrqvrp-------------------------pscpvpfpet 36

Qy    162 FDGNPDMLAPFMAQCQIFMEKSTRDFSVDRVRVCFVTSMMTGRAARWASAKLERSHYLMH 221
          |:|    |  |: |   :|  :   |  | ::| |: |::|| |  |    :|    ::
Db     37 fngessrlpefivqtasymlvnenrfcndamkvaflislltgeaeewvvpyiemdspilg 96

Qy    222 NYPAFMMEMKHVF-----EDPQRRE 241
          :| ||: |||  |     ||    |
Db     97 dyrafldemkqcfgwdddeddddee 121


RESULT   8
AAY26054
ID   AAY26054 standard; Protein; 110 AA.
XX
AC   AAY26054;
XX
DT   28-SEP-1999  (first entry)
XX
DE   Human Zalpha1 epitope-bearing protein fragment 3.
XX
KW   Alpha helical protein-1; Fragile-X syndrome; cardiovascular system;
KW   connective tissue; abnormal proliferative disorder; cancer; skin tone;
KW   epidermal system; cosmetic improvement; skin tone; elasticity; Zalpha1;
KW   epitope.
XX
OS   Homo sapiens.
XX
PN   WO9929720-A2.
XX
PD   17-JUN-1999.
XX
PF   10-DEC-1998;   98WO-US26273.
XX
PR   10-DEC-1997;   97US-0987926.
XX
PA   (ZYMO ) ZYMOGENETICS INC.
XX
PI   Conklin DC,  Lok S,  Parrish J;
XX
DR   WPI; 1999-385572/32.
XX
PT   Mammalian alpha helical protein-1, designated Zalpha1
XX
PS   Claim 10; Page 70; 73pp; English.
XX
CC   The present sequence is an epitope-bearing protein derived from
CC   human alpha helical protein-1, Zalpha1. The protein comprises
CC   helices A, B, C and D of mature Zalpha1. It is used to raise specific
CC   antibodies which can be used for detection and purification of Zalpha1.
CC   The Zalpha1 protein may be useful in the treatment of Fragile-X
CC   syndrome and abnormal proliferative disorders e.g. cancer. It can also
CC   be used for the growth, differentiation, maintenance and survival of
CC   connective tissues, particularly cardiovascular and epidermal systems
CC   and in imparting cosmetic improvements to normal connective tissues such
CC   as enhancement of skin tone and elasticity.
XX
SQ   Sequence   110 AA;


  Query Match             7.0%;  Score 149.5;  DB 20;  Length 110;
  Best Local Similarity   27.3%;  Pred. No. 0.00014;
  Matches   36;  Conservative   18;  Mismatches   53;  Indels   25;  Gaps    1;

Qy    103 NNNLQSQVQKLTEENTTLREQVEPTPEDEDDDIELRGAAAAAAPPPPIEEECPEDLPEKF 162
          |: |  |:: |  |  :|  || |                              ||   || |
Db      1 nsqlmeqlrllvcerasllrqvrp-------------------------pscpvpfpetf 35

Qy    163 DGNPDMLAPFMAQCQIFMEKSTRDFSVDRVRVCFVTSMMTGRAARWASAKLERSHYLMHN 222
          :|    |  |: |   :|  :   |  | ::| |: |::|| |  |    :|    :: :
Db     36 ngessrlpefivqtasymlvnenrfcndamkvaflislltgeaeewvvpyiemdspilgd 95

Qy    223 YPAFMMEMKHVF 234
          | ||: |||  |
Db     96 yrafldemkqcf 107


RESULT   9
AAY94674
ID   AAY94674 standard; Protein; 168 AA.
XX
AC   AAY94674;
XX
DT   01-DEC-2000  (first entry)
XX
DE   Human zsig83 mature protein sequence.
XX
KW   Alpha-helical protein; zsig83; cell growth; differentiation; cancer;
KW   proliferation; chromosome 22q13.1-q13.2; cytostatic; vulneray;
KW   degenerative condition; metastasis; wound healing.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Domain          22..38
FT                   /note= "SH3-binding domain"
FT   Region          41..55
FT                   /label= Helix_A
FT   Region          60..65
FT                   /note= "Hydrophilic region"
FT   Region          63..81
FT                   /label= Helix_B
FT   Region          74..79
FT                   /note= "Hydrophilic region"
FT   Region          97..105
FT                   /label= Helix_C
FT   Region          103..108
FT                   /note= "Hydrophilic region"
FT   Region          110..119
FT                   /label= Helix_D
FT   Region          126..131
FT                   /note= "Hydrophilic region"
FT   Region          128..138
FT                   /label= Helix_E
FT   Region          162..167
FT                   /note= "Hydrophilic region"
XX
PN   WO200050594-A2.
XX
PD   31-AUG-2000.
XX
PF   25-FEB-2000; 2000WO-US04816.
XX
PR   26-FEB-1999;   99US-0259131.
XX
PA   (ZYMO ) ZYMOGENETICS INC.
XX
PI   Presnell SR;
XX
DR   WPI; 2000-572091/53.
DR   N-PSDB; AAA28032.
XX
PT   Alpha-helical protein zsig83, its antibodies and the polynucleotide
PT   encoding the protein useful for treating disorders associated with
PT   abnormal cell growth e.g. cancer and agonists useful for treating
PT   wounds  -
XX
PS   Claim 4; Page 75; 83pp; English.
XX
CC   This invention relates to a novel human alpha-helical protein designated
CC   zsig83. Zsig83 plays a role in the process of cell growth,
CC   differentiation, or proliferation. The zsig83 gene is located on
CC   chromosome 22 at position 22q13.1-q13.2. Included in the invention are
CC   polynucleotide sequences encoding the zsig83 protein, expression vectors
CC   containing the zsig83 DNA sequence, a cultured cell containing the
```

CC expression vector, and antibodies specific to the zsig83 protein. The
CC zsig83 protein contains 5 alpha helix regions (represented by sequences
CC AAY94677-Y94681) and also contains epitope bearing regions (represented
CC by sequences AAY94688-Y94698) to which the antibodies are directed. The
CC protein exhibits cytostatic and vulneray activity. The zsig83 protein
CC and nucleotide sequences and antibodies are used for treating disorders
CC associated with abnormal cell growth e.g. cancer, degenerative
CC conditions and metastasis. The zsig83 protein and its agonists or
CC antagonists are useful for promoting wound healing. The zsig83 DNA
CC sequence can be used to identify defective zsig83 genes and may therefore
CC be used as a diagnostic indicator of cancer.
CC The present sequence represents the mature human zsig83 protein.
XX
SQ Sequence 168 AA;

Query Match 7.0%; Score 149.5; DB 21; Length 168;
Best Local Similarity 33.7%; Pred. No. 0.00024;
Matches 32; Conservative 17; Mismatches 45; Indels 1; Gaps 1;

Qy 155 PEDLPEKFDGNPDMLAPFMAQCQIFMEKSTRDFSVDRVRVCFVTSMMTGRAARWASAKLE 214
Db 27 ptslpepfsgdpgrlagflmqmdrfmifqasrfpgeaervaflvsrltgeaekwaiphmq 86

Qy 215 RSHYLMHNYPAFMMEMKHVFEDPQRREVAKRKIRR 249
Db 87 pdsplrnnyqgflaelrrtyksp-lrharraqirk 120

RESULT 10
AAY94673
ID AAY94673 standard; Protein; 184 AA.
XX
AC AAY94673;
XX
DT 01-DEC-2000 (first entry)
XX
DE Human zsig83 protein sequence.
XX
KW Alpha-helical protein; zsig83; cell growth; differentiation; cancer;
KW proliferation; chromosome 22q13.1-q13.2; cytostatic; vulneray;
KW degenerative condition; metastasis; wound healing.
XX
OS Homo sapiens.
XX

FH	Key	Location/Qualifiers
FT	Peptide	1..16
FT		/note= "Putative signal sequence"
FT	Protein	17..184
FT		/label= zsig83
FT	Domain	38..54
FT		/note= "SH3-binding domain"
FT	Region	57..71
FT		/label= Helix_A
FT	Region	79..97
FT		/label= Helix_B
FT	Region	76..81
FT		/note= "Hydrophilic region"
FT	Region	90..95
FT		/note= "Hydrophilic region"
FT	Region	113..121
FT		/label= Helix_C
FT	Region	119..124
FT		/note= "Hydrophilic region"
FT	Region	126..135
FT		/label= Helix_D
FT	Region	144..154
FT		/label= Helix_E
FT	Region	142..147
FT		/note= "Hydrophilic region"
FT	Region	178..183
FT		/note= "Hydrophilic region"

XX
PN WO200050594-A2.
XX
PD 31-AUG-2000.
XX
PF 25-FEB-2000; 2000WO-US04816.
XX
PR 26-FEB-1999; 99US-0259131.
XX
PA (ZYMO) ZYMOGENETICS INC.
XX
PI Presnell SR;
XX
DR WPI; 2000-572091/53.
DR N-PSDB; AAA28032, AAA28037.
XX
PT Alpha-helical protein zsig83, its antibodies and the polynucleotide
PT encoding the protein useful for treating disorders associated with
PT abnormal cell growth e.g. cancer and agonists useful for treating
PT wounds -
XX
PS Claim 1; Page 73-74; 83pp; English.
XX
CC This invention relates to a novel human alpha-helical protein designated
CC zsig83. Zsig83 plays a role in the process of cell growth,
CC differentiation, or proliferation. The zsig83 gene is located on
CC chromosome 22 at position 22q13.1-q13.2. Included in the invention are
CC polynucleotide sequences encoding the zsig83 protein, expression vectors
CC containing the zsig83 DNA sequence, a cultured cell containing the
CC expression vector, and antibodies specific to the zsig83 protein. The
CC zsig83 protein contains 5 alpha helix regions (represented by sequences
CC AAY94677-Y94681) and also contains epitope bearing regions (represented
CC by sequences AAY94688-Y94698) to which the antibodies are directed. The
CC protein exhibits cytostatic and vulneray activity. The zsig83 protein
CC and nucleotide sequences and antibodies are used for treating disorders
CC associated with abnormal cell growth e.g. cancer, degenerative
CC conditions and metastasis. The zsig83 protein and its agonists or
CC antagonists are useful for promoting wound healing. The zsig83 DNA
CC sequence can be used to identify defective zsig83 genes and may therefore
CC be used as a diagnostic indicator of cancer.
CC The present sequence represents the human zsig83 protein.
XX
SQ Sequence 184 AA;

Query Match 7.0%; Score 149.5; DB 21; Length 184;
Best Local Similarity 33.7%; Pred. No. 0.00027;
Matches 32; Conservative 17; Mismatches 45; Indels 1; Gaps 1;

Qy 155 PEDLPEKFDGNPDMLAPFMAQCQIFMEKSTRDFSVDRVRVCFVTSMMTGRAARWASAKLE 214
Db 43 ptslpepfsgdpgrlagflmqmdrfmifqasrfpgeaervaflvsrltgeaekwaiphmq 102

Qy 215 RSHYLMHNYPAFMMEMKHVFEDPQRREVAKRKIRR 249
Db 103 pdsplrnnyqgflaelrrtyksp-lrharraqirk 136

RESULT 11
ABG12205
ID ABG12205 standard; Protein; 142 AA.
XX
AC ABG12205;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #12196.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.

```
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   N-PSDB; AAS76392.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
XX
PS   Claim 20; SEQ ID No 42564; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human
CC   diagnostic amino acid sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   142 AA;


  Query Match             6.5%;  Score 139.5;  DB 22;  Length 142;
  Best Local Similarity   31.4%;  Pred. No. 0.0012;
  Matches   38;  Conservative   18;  Mismatches   54;  Indels   11;  Gaps    3;

Qy     129 EDEDDDIELRGAAAA-------AAPPPPIEEEC--PEDLPEKFDGNPDMLAPFMAQCQIF 179
           |  |   :||    |       | |  |       |   || |||: | |  |: |   :
Db      20 easdrsedlrwtvcaqliqallalpirpatrrwrnpipfpetfdgdtdrlpefivqtcsy 79

Qy     180 MEKSTRDFSVDRVRVCFVTSMMTGRAARWASAKLERSHYLMHNYPAFMMEMKHVF--EDP 237
           |      || | ::| |: : :|| | :|    :::   |:::|  |: ||| ||  ||
Db      80 mivdentfssdalkvtflitrltgpalqwvipyikkespllndyrgflaemkrvfgwedd 139

Qy     238 Q 238
           :
Db     140 e 140


RESULT  12
AAB60475
ID   AAB60475 standard; Protein; 113 AA.
XX
AC   AAB60475;
XX
DT   24-APR-2001  (first entry)
XX
DE   Human cell cycle and proliferation protein CCYPR-23, SEQ ID NO:23.
XX
KW   Cell cycle and proliferation protein; CCYPR; human; agonist;
KW   antagonist; gene therapy; detection; gene therapy;
KW   transgenic animal disease model; immune disorder;
KW   developmental disorder; cell signalling disorder;
KW   cell proliferative disorder; cancer; tumour; anaemia; epilepsy;
KW   arteriosclerosis; asthma; allergy; diabetes mellitus;
KW   menstrual cycle disorder; bacterial infection.
XX
OS   Homo sapiens.
XX
PN   WO200107471-A2.
XX
PD   01-FEB-2001.
XX
PF   21-JUL-2000; 2000WO-US19948.
XX
PR   21-JUL-1999;   99US-0145075.
PR   08-SEP-1999;   99US-0153129.
PR   10-NOV-1999;   99US-0164647.
XX
PA   (INCY-) INCYTE GENOMICS INC.
XX
PI   Hillman JL,  Lal P,  Tang YT,  Yue H,  Au-Young J,  Bandman O;
PI   Azimzai Y,  Yang J,  Lu DAM,  Baughn MR,  Patterson C,  Shah P;
XX
DR   WPI; 2001-112727/12.
DR   N-PSDB; AAF59612.
XX
PT   Human cell cycle and proliferation proteins and polynucleotides are
PT   used to treat, diagnose and prevent immune, developmental and cell
PT   signaling disorders and cell proliferative disorders including cancer -
XX
PS   Claim 1; Page 131; 205pp; English.
XX
CC   Sequences AAB60453-AAB60506 represent 54 human cell cycle and
CC   proliferation proteins (CCYPR), which are encoded by AAF59590-AAF59643.
CC   CCYPR and agonists of CCYPR are used to treat diseases or conditions
CC   associated with decreased expression of functional CCYPR, while CCYPR
CC   antagonists are used to treat diseases or conditions associated with
CC   overexpression of functional CCYPR. Monoclonal or polyclonal antibodies
CC   to CCYPR may be used in enzyme-linked immunosorbent assays (ELISA) or
CC   radioimmunoasays to detect CCYPR. CCYPR itself may be used to detect
CC   compounds e.g., antibodies, oligonucleotides and proteins (receptors)
CC   that specifically bind to CCYPR, and in drug screening methods to
CC   identify compounds that modulate the activity of CCYPR. CCYPR
CC   nucleotides can be used to generate transgenic animal models of human
CC   disease, and can be used in gene therapy in target cells with genetic
CC   abnormalities with respect to the expression of CCYPR for the
CC   treatment or prevention of a disorder associated with CCYPR.
CC   Diseases which can be diagnosed, treated and prevented using CCYPR
CC   proteins, nucleic acids, agonists or antagonists include immune,
CC   developmental and cell signalling disorders, and cell proliferative
CC   disorders including cancer. Specific examples of these disorders
CC   include anaemia, epilepsy, arteriosclerosis, asthma, cancer, allergies,
CC   diabetes mellitus, disorders of the menstrual cycle and infections
CC   caused by bacteria.
XX
SQ   Sequence   113 AA;


  Query Match             6.5%;  Score 139;  DB 22;  Length 113;
  Best Local Similarity   35.0%;  Pred. No. 0.001;
  Matches   28;  Conservative   16;  Mismatches   36;  Indels    0;  Gaps    0;

Qy     155 PEDLPEKFDGNPDMLAPFMAQCQIFMEKSTRDFSVDRVRVCFVTSMMTGRAARWASAKLE 214
           |   || |||: | |  |: |   :|      || | ::| |: : :|| | :|    ::
Db      26 pipfpetfdgdtdrlpefivqtgsymfvdentfssdalkvtflitrltgpalqwvipyik 85

Qy     215 RSHYLMHNYPAFMMEMKHVF 234
           :   |:::|  |: ||| ||
Db      86 kespllndyrgflaemkrvf 105
```

RESULT 13
AAY59927
ID AAY59927 standard; Protein; 144 AA.
XX
AC AAY59927;
XX
DT 28-JAN-2000 (first entry)
XX
DE Human myometrium tumour EST encoded protein 7.
XX
KW Myometrium; tumour; human; expressed sequence tag; EST; uterine myoma;
KW treatment; carcinoma; cancer; gene therapy.
XX
OS Homo sapiens.
XX
PN DE19817947-A1.
XX
PD 28-OCT-1999.
XX
PF 17-APR-1998; 98DE-1017947.
XX
PR 17-APR-1998; 98DE-1017947.
XX
PA (META-) METAGEN GES GENOMFORSCHUNG MBH.
XX
PI Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E;
XX
DR WPI; 1999-602380/52.
DR N-PSDB; AAZ41965.
XX
PT New nucleic acid sequences expressed in uterine myoma, and derived
PT polypeptides, for treatment of uterine carcinoma and identification of
PT therapeutic agents -
XX
PS Claim 23; Page 68; 86pp; German.
XX
CC This invention describes novel polypeptide sequences (I), fragments of
CC (I) fragments and their encoding nucleic acids (II) which are highly
CC expressed in human uterine myoma. (II) are used for recombinant
CC expression of (I) and to isolate complete genes. (I) are used to
CC identify agents suitable for treatment of uterine carcinoma, to directly
CC treat this form of cancer (including expression from gene therapy
CC vectors) and are used in a preparation for cancer treatment. (I) is also
CC used for the generation of specific antibodies. (II) are identified by
CC assembling ESTs (expressed sequence tags) from a particular tissue type
CC before comparison of expression patterns. This allows a significantly
CC longer fragment of the gene to be revealed and therefore reduces the
CC number of failures associated with the fact that ESTs from different
CC libraries may represent different parts of the same unknown gene,
CC distorting the estimated frequency of occurrence in a particular tissue.
CC AAY59921-Y59940 represent protein fragments encoded by the human
CC myometrium tumour cDNA library derived EST fragments represented in
CC AAZ41950-Z41980.
XX
SQ Sequence 144 AA;

Query Match 6.4%; Score 137; DB 20; Length 144;
Best Local Similarity 29.4%; Pred. No. 0.002;
Matches 37; Conservative 19; Mismatches 58; Indels 12; Gaps 3;

```
Qy    121 REQVEPTPE-DEDDDIELRGA---------AAAAAPPPPIEEEC--PEDLPEKFDGNPDM 168
          |  ::   |   : :  | ||          |  | |  |       |   || |||: |
Db     11 rrrlsadphatqrnsaeargtmdgrvqlmkallagplrpaarrwrnpipfpetfdgdtdr 70

Qy    169 LAPFMAQCQIFMEKSTRDFSVDRVRVCFVTSMMTGRAARWASAKLERSHYLMHNYPAFMM 228
          |  |: |   :|      || | ::| |: : :|| | :|    : :   |:::|  |:
Db     71 lpefivqtcsymfvdentfsndalkvtflitrltgpalqwvipyirkespllndyrgfla 130

Qy    229 EMKHVF 234
          ||| ||
Db    131 emkrvf 136
```

RESULT 14
AAW48895
ID AAW48895 standard; Protein; 1230 AA.
XX
AC AAW48895;
XX
DT 13-OCT-1998 (first entry)
XX
DE Candida albicans CST20 protein.
XX
KW CST20; protein kinase; Ste20p/p65PAK family; screening; virulence;
KW hyphal formation; pathogenic fungi; inhibitor; inflammation;
KW antimycotic.
XX
OS Candida albicans.
XX
PN WO9818927-A1.
XX
PD 07-MAY-1998.
XX
PF 29-OCT-1997; 97WO-CA00809.
XX
PR 30-OCT-1996; 96US-0029458.
XX
PA (CANA) NAT RES COUNCIL CANADA.
XX
PI Leberer E, Thomas DY;
XX
DR WPI; 1998-272222/24.
DR N-PSDB; AAV32553.
XX
PT In vitro screening test for agents that inhibit Candida genes
PT involved in virulence - and transition to hyphal form, potentially
PT useful as antimycotic agents
XX
PS Disclosure; Fig 3; 79pp; English.
XX
CC The sequence is that of the CST20 protein which can be used
CC in the development of an in vitro screening test for compounds
CC that inhibit biological activity of the protein and a system for
CC measuring its activity. The protein is involved in virulence and
CC hyphal formation. Inhibitors are potentially useful for rendering
CC pathogenic fungi (any species in which hyphal induction by kinase
CC occurs) avirulent and/or to treat inflammation.
XX
SQ Sequence 1230 AA;

Query Match 6.3%; Score 135.5; DB 19; Length 1230;
Best Local Similarity 22.5%; Pred. No. 0.043;
Matches 67; Conservative 54; Mismatches 114; Indels 63; Gaps 15;

```
Qy     17 GGQDPGLHPHR-SEATAGRSPPTPTVTLGPDCPPPPPPPPPNNNNNNNSKHTGHKSA--- 72
          ||::  |   | :|  | |:|| |   | |  || || || |  :   |:    ::|
Db    770 ggennalpkqrinefkahrapppp--plappappvppappanllseqtseipqqrtaplq 827

Qy     73 ------CVPNMTERRRDELSEEINNLREKVMKQSEENNNLQSQVQKLTEENTTLREQVEP 126
                |: | :: :  |    ||||  :: ||       :|:| |:|     | |
Db    828 aladvtaptniyeiqqtkyqeaqqklrekkarelee-------iqrlreknerqnrqqet 880

Qy    127 TPEDEDDDIELRGAAAAAAPPPPIEEECPEDLPEKFDGNPD-MLAPFMAQ---------- 175
          : |         :  ||| |:  : |   |    |  |   |  :||
Db    881 gqnnad----tasggsniappvpvpnkkp---psgsgggrdakqaaliaqkkreekkrkn 933

Qy    176 CQIFMEKST-------RDFSVDRVRVCFVTSMMTGRAARWASAKLERSHYLMHNYPAFMM 228
          ||  : |         :  || |::           : ||  :  :| :        :
Db    934 lqiiaklkticnpgdpnelyvdlvki----------gqgasggvflahdvrdksnivai 982

Qy    229 EMKHVFEDPQRREVAKRKIRRLRQGM-GSVIDYSNAFQMIAQDLDWNEPALIDQYHEG 285
          :  :: : | ::|:   :|  ::  :  ::::: ::: ::  || |    :| :| ||
Db    983 kqmnleqqp-kkeliineilvmkgslhpnivnfidsy-llkgdl-w----vimeymeg 1033
```

```
RESULT 15
AAY20666
ID   AAY20666 standard; Protein; 557 AA.
XX
AC   AAY20666;
XX
DT   22-JUL-1999  (first entry)
XX
DE   Human neurofilament-M wild type protein fragment 8.
XX
KW   Human; beta-amyloid precursor protein; beta-APP; diagnosis; cancer;
KW   frameshift mutation; age-related disease; neurodegenerative disorder;
KW   Alzheimer's disease; Down's syndrome; myotonic dystrophy; neuronal;
KW   Huntington's disease; multiple sclerosis; alcoholic liver disease;
KW   diabetes mellitus type II; microtubule associated protein; Tau; Big Tau;
KW   ubiquitin B; apolipoprotein E; MAP2; neurofilament-L; neurofilament-M;
KW   neurofilament-F; presenilin I; presenilin II; cellular tumour antigen;
KW   glial fibrillary acidic protein; GFAP; p53; semaphorin III; HUPF-1;
KW   bcl-2; B-cell leukemia/lymphoma 2 proto-oncogene; HMGP-C; NSP-A;
KW   high mobility group protein-C; neuroendocrine specific protein A.
XX
OS   Homo sapiens.
XX
PN   WO9845322-A2.
XX
PD   15-OCT-1998.
XX
PF   02-APR-1998;   98WO-IB00705.
XX
PR   10-APR-1997;   97US-0043163.
XX
PA   (UYUT-) RIJKSUNIV UTRECHT.
PA   (ROYA-) ROYAL NETHERLANDS ACAD ARTS & SCI.
PA   (UYRO-) UNIV ROTTERDAM ERASMUS.
XX
PI   Burbach JPH,  Grosveld FG,  Van Leeuwen FW;
XX
DR   WPI; 1998-609901/51.
DR   N-PSDB; AAX75759.
XX
PT   Diagnosing disease by detecting frameshift mutations in RNA or
PT   corresponding protein mutations - used to diagnose cancer and
PT   neurological diseases, particularly Alzheimer's disease, and also
PT   for treatment and prevention with specific ribozymes or wild-type
PT   RNA
XX
PS   Disclosure; Figure 8; 258pp; English.
XX
CC   This invention describes a novel method for the diagnosis of a disease
CC   caused by, or associated with, an RNA molecule that has a frameshift
CC   mutation. The method is used to diagnose age-related diseases, especially
CC   cancer and a wide range of neurodegenerative disorders (e.g. Alzheimer's
CC   disease, Down's syndrome, myotonic dystrophy, Huntington's disease,
CC   multiple sclerosis, alcoholic liver disease, diabetes mellitus type II
CC   and many others listed) or susceptibility to these disorders. The method
CC   allows a definitive diagnosis of Alzheimer's disease in living patients,
CC   at an early stage. It is based on the observation that disease may be
CC   caused by mutations in RNA rather than DNA. The invention describes the
CC   used of neuronal system RNA molecules, specifically proteins including
CC   beta-amyloid precursor protein (beta-APP), the microtubule associated
CC   proteins Tau and Big Tau, ubiquitin B, apolipoprotein E, microtubule
CC   associated protein 2 (MAP2), neurofilament-L, neurofilament-M,
CC   neurofilament-F, presenilin I, presenilin II, glial fibrillary acidic
CC   protein (GFAP), the cellular tumour antigen p53, B-cell leukemia/lymphoma
CC   2 (bcl-2) proto-oncogene, semaphorin III, HUPF-1, high mobility group
CC   protein-C (HMGP-C) and neuroendocrine specific protein A.
XX
SQ   Sequence   557 AA;

  Query Match             6.2%;  Score 132;  DB 19;  Length 557;
  Best Local Similarity   23.7%;  Pred. No. 0.03;
  Matches   74;  Conservative   47;  Mismatches  125;  Indels   66;  Gaps   17;

Qy      13 RSGRGGQDP---------GLHPHRSEATAGRSPPTPTVTLGPD-CP--PPPPPPPPNNNN 60
           |   ||  |         || |  : | :  | | |   | || ||  ||| ||||
Db       5 rgpgggagpraeksrfpcglgprgagaespfsrppp---llpdrcpqsppprpppp---- 57

Qy      61 NNNSKHTGHKSACVPNMTERRRDELSEEINNLREKVMKQSEENNNL--QSQVQKLTEENT 118
                 |  :| :  :   |::     :   | |  | |   ::|   |  :::||  :
Db      58 -rggpwpgtsaaainglrrgrqn----avtatrpkaskmsytldslgnpsayrrvtetrs 112

Qy     119 TLREQVEPTPEDEDDDIELRGAAAAAAPPPPIEEECPEDLPEKFDGNPDMLAPFMAQCQI 178
           :   :|  :|          |  : :   |  :                  |||| :|
Db     113 sf-srvsgspss-----gfrsqswsrgspstvsssykr----------smlaprlayssa 156

Qy     179 FME--KSTRDFSVDRVRVCFVTSMMTGRAARWASAKLERSHYL-----MHNYPAFMMEMK 231
            :  :|: |||         :|:: | :      || ||:       :::  |  :|
Db     157 mlssaessldfsqs------ssllnggsgpggdyklsrsnekeqlqglndrfagyiekv 209

Qy     232 HVFEDPQRREVAKRKIRRLRQGMGSVIDYSNAFQMIAQDLDWNEPALIDQYHEGL----- 286
           |  |  | :|: : :|: |||   |     :|:    ::|      |    ||
Db     210 hyleq-qnkei-eaeiqalrqkqashaqlgdaydqeirel---ratlemvnhekaqvqld 264

Qy     287 SDHIQEELSHLE 298
           |||::|::  |:
Db     265 sdhleedihrlk 276


Search completed: October  4, 2002, 15:39:48
Job time: 90 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: October 4, 2002, 15:38:18 ; Search time 13.41 Seconds
(without alignments)
1157.833 Million cell updates/sec

Title: US-09-631-863A-2
Perfect score: 2139
Sequence: 1 MRNKRVLKTKKRRSGRGGQD.....DNCPAKASKSSPAGNSPAPL 401

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	159	7.4	146	1	LDO1_HUMAN	O95751 homo sapien
2	156	7.3	520	1	CET1_CANAL	O93803 candida alb
3	135	6.3	537	1	GAG_MLVRD	P11269 radiation m
4	131.5	6.1	538	1	GAG_MLVFF	P26806 friend muri
5	131.5	6.1	1230	1	ST20_CANAL	Q92212 candida alb
6	130	6.1	537	1	GAG_MLVBM	P29167 murine leuk
7	129	6.0	872	1	S3B2_HUMAN	Q13435 homo sapien
8	122.5	5.7	536	1	GAG_MLVDE	P29168 murine leuk
9	122.5	5.7	538	1	GAG_MLVFP	P26805 friend muri
10	121.5	5.7	1585	1	P3K3_DICDI	P54675 dictyosteli
11	120.5	5.6	1157	1	Y182_HUMAN	Q14687 homo sapien
12	120.5	5.6	1816	1	AF6_HUMAN	P55196 homo sapien
13	120.5	5.6	1982	1	CHDM_DROME	O97159 drosophila
14	120	5.6	745	1	HGL2_ARATH	P46607 arabidopsis
15	120	5.6	1905	1	TAGB_DICDI	P54683 dictyosteli
16	120	5.6	1937	1	MYH8_HUMAN	P13535 homo sapien
17	119.5	5.6	1939	1	MYH1_HUMAN	P12882 homo sapien
18	119	5.6	465	1	MYH6_RABIT	P04460 oryctolagus
19	119	5.6	505	1	WASL_HUMAN	O00401 homo sapien
20	118.5	5.5	428	1	INVO_CEBAL	P24709 cebus albif
21	118.5	5.5	2564	1	SPCQ_HUMAN	Q9h254 homo sapien
22	118	5.5	2175	1	HMCU_DROME	P10180 drosophila
23	117.5	5.5	501	1	WASL_RAT	O08816 rattus norv
24	116.5	5.4	281	1	CEBE_RAT	P56261 rattus norv
25	116.5	5.4	957	1	NED4_MOUSE	P46935 mus musculu
26	116.5	5.4	2142	1	BAT2_HUMAN	P48634 homo sapien
27	116	5.4	537	1	GAG_MLVAV	P03336 akv murine
28	116	5.4	3680	1	DMD_CANFA	O97592 canis famil
29	115.5	5.4	536	1	GAG_MLVCB	P27460 cas-br-e mu
30	115	5.4	1227	1	B3A3_MOUSE	P16283 mus musculu
31	115	5.4	1332	1	SPT7_YEAST	P35177 saccharomyc
32	115	5.4	1935	1	MYH7_HUMAN	P12883 homo sapien
33	115	5.4	1935	1	MYH7_PIG	P79293 sus scrofa
34	115	5.4	1939	1	MYH4_HUMAN	Q9y623 homo sapien
35	114.5	5.4	622	1	LAM0_DROME	P08928 drosophila
36	114.5	5.4	633	1	LA17_YEAST	Q12446 saccharomyc
37	114.5	5.4	1941	1	MYH2_HUMAN	Q9ukx2 homo sapien
38	114	5.3	640	1	ELL2_HUMAN	O00472 homo sapien
39	113.5	5.3	281	1	CEBE_HUMAN	Q15744 homo sapien
40	113	5.3	1727	1	ALM1_SCHPO	Q9utk5 schizosacch
41	113	5.3	1934	1	MYH7_MESAU	P13540 mesocricetu
42	112.5	5.3	2842	1	APC_RAT	P70478 rattus norv
43	112	5.2	502	1	WASP_HUMAN	P42768 homo sapien
44	112	5.2	1223	1	YWR1_CAEEL	Q10925 caenorhabdi
45	111.5	5.2	1232	1	B3A3_HUMAN	P48751 homo sapien

ALIGNMENTS

RESULT 1
LDO1_HUMAN
ID LDO1_HUMAN STANDARD; PRT; 146 AA.
AC O95751;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE LDOC1 protein (Leucine zipper protein down-regulated in cancer cells).
GN LDOC1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Fetal brain;
RX MEDLINE=99330357; PubMed=10403563;
RA Nagasaki K., Manabe T., Hanzawa H., Maass N., Tsukada T.,
RA Yamaguchi K.;
RT "Identification of a novel gene, LDOC1, down-regulated in cancer cell
RT lines.";
RL Cancer Lett. 140:227-234(1999).
CC -!- FUNCTION: MAY HAVE AN IMPORTANT ROLE IN THE DEVELOPMENT AND/OR
CC PROGRESSION OF SOME CANCERS.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- TISSUE SPECIFICITY: UBIQUITOUSLY EXPRESSED WITH HIGH LEVELS IN
CC BRAIN ANT THYROID AND LOW EXPRESSION IN PLACENTA, LIVER AND
CC LEUCOCYTES. EXPRESSED AS WELL IN SIX OF THE SEVEN HUMAN BREAST
CC CANCER CELL LINES EXAMINED.
CC ----------

CC ----------
DR EMBL; AB019527; BAA34364.1; -.
KW Nuclear protein.
FT DOMAIN 132 143 ASP/GLU-RICH (HIGHLY ACIDIC).
SQ SEQUENCE 146 AA; 16968 MW; 3D02813B2DE52DBE CRC64;

Query Match 7.4%; Score 159; DB 1; Length 146;
Best Local Similarity 27.4%; Pred. No. 0.0007;
Matches 45; Conservative 21; Mismatches 68; Indels 30; Gaps 2;

```
Qy      83 DELSEEINNLREKVMKQSEENNNLQSQVQKLTEENTTLREQVEPTPEDEDDDIELRGAAA 142
           |||   :: |  :    | ||: |  |:: |  |  :|  || |
Db       3 DELVLLLHALLMRHRALSIENSQLMEQLRLLVCERASLLRQVRP---------------- 46

Qy     143 AAAPPPPIEEECPEDLPEKFDGNPDMLAPFMAQCQIFMEKSTRDFSVDRVRVCFVTSMMT 202
                   ||   || |:|    |  |: |   :|  :   |  | ::| |: |::|
Db      47 ---------PSCPVPFPETFNGESSRLPEFIVQTASYMLVNENRFCNDAMKVAFLISLLT 97
```

```
Qy     203 GRAARWASAKLERSHYLMHNYPAFMMEMKHVF-----EDPQRRE 241
Db      98 GEAEEWVVPYIEMDSPILGDYRAFLDEMKQCFGWDDDEDDDDEE 141


RESULT   2
CET1_CANAL
ID   CET1_CANAL     STANDARD;      PRT;   520 AA.
AC   O93803;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   mRNA capping enzyme beta subunit (Polynucleotide 5'-triphosphatase)
DE   (EC 3.1.3.33) (mRNA 5'-triphosphatase) (TPase).
GN   CET1.
OS   Candida albicans (Yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX   NCBI_TaxID=5476;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=IFO 1060;
RX   MEDLINE=98427288; PubMed=9755857;
RA   Yamada-Okabe T., Mio T., Matsui M., Kashima Y., Arisawa M.,
RA   Yamada-Okabe H.;
RT   "Isolation and characterization of the Candida albicans gene for mRNA
RT   5'-triphosphatase: association of mRNA 5'-triphosphatase and mRNA 5'-
RT   guanylyltransferase activities is essential for the function of mRNA
RT   5'-capping enzyme in vivo.";
RL   FEBS Lett. 435:49-54(1998).
CC   -!- FUNCTION: FIRST STEP OF M-RNA CAPPING. CONVERTS THE 5'-
CC       TRIPHOSPHATE END OF A NASCENT MRNA CHAIN INTO A DIPHOSPHATE END.
CC   -!- CATALYTIC ACTIVITY: 5'-phosphopolynucleotide + H(2)O =
CC       polynucleotide + phosphate.
CC   -!- COFACTOR: REQUIRES DIVALENT IONS (BY SIMILARITY).
CC   -!- SUBUNIT: THE M-RNA CAPPING ENZYME IS COMPOSED OF TWO SEPARATE
CC       CHAINS ALPHA AND BETA, RESPECTIVELY A MRNA GUANYLYLTRANSFERASE AND
CC       AN RNA 5'-TRIPHOSPHATASE.
CC   -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE FUNGAL TPASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AB016242; BAA33965.1; -.
DR   InterPro; IPR004206; mRNA_triPase.
DR   Pfam; PF02940; mRNA_triPase; 1.
KW   Hydrolase; mRNA processing; mRNA capping; Nuclear protein.
FT   DOMAIN       92     95       POLY-SER.
SQ   SEQUENCE   520 AA;  58791 MW;  5C0690F547E13EEC CRC64;


  Query Match             7.3%;  Score 156;  DB 1;  Length 520;
  Best Local Similarity   21.6%;  Pred. No. 0.0048;
  Matches   78;  Conservative   59;  Mismatches  118;  Indels  106;  Gaps   18;

Qy       9 TKKRRSGRGGQDPGLHPHRS-----EATAGRSPPT--PTVTLGPDCPPP---------- 50
Db      24 TKIIKSPTAYHKPSVHERHSITSMLNDTPSDSTPTKKPEPTISPEFRKPSISSLTSPSVA 83

Qy      51 --PPPPPPNNNNNNNSKHTGHKSACVPNMTERR-------------RDELSEEINNLRE 94
Db      84 HKPPPLPPSSSSVGSSEHSSARSS--PAITKRNSIANIIDAYEEPATKTEKKAELNSPK- 140

Qy      95 KVMKQS------EENNNLQSQVQKLTEENTTLREQVEPTP--EDEDDDI-ELRGAAAAAA 145
Db     141 --INQSTPVPKLEEHENDTNKVEKVVDSAPEPKPKKEPQPVFDDQDDDLTKIKKLKQSKK 198

Qy     146 P----PPPI-----------EEECPEDLPEKFDGNPDMLAPFMAQCQIFMEKSTRDFSVD 190
Db     199 PRRYETPPIWAQRWVPPNRQKEETNVD-----DGNEAITR--LSEKPVFDYTTTRSVDLE 251

Qy     191 RVRVCFVTSM-----MTGRAARWASAKLERSHYLMHNYPAFMMEMKHVFEDPQRREVAKR 245
Db     252 ----CSITGMIPPSSITRKIAEWVYANFSN-----------------VEEKSKRNV--- 286

Qy     246 KIRRLRQGMGSVIDYSNAFQMIAQDLDWNEPALIDQYHEGLSDHIQEELSHLEVAKSLSA 305
Db     287 ---ELELKFGKIIDKRSGNRI---DLNVVTECIFTDHSSVFFDMQVEEVAWKEITKFLDE 340

Qy     306 L 306
Db     341 L 341


RESULT   3
GAG_MLVRD
ID   GAG_MLVRD      STANDARD;      PRT;   537 AA.
AC   P11269;
DT   01-JUL-1989 (Rel. 11, Created)
DT   01-JUL-1989 (Rel. 11, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   GAG polyprotein [Contains: Core protein P15; Inner coat protein P12;
DE   Core shell protein P30; Nucleoprotein P10].
GN   GAG.
OS   Radiation murine leukemia virus.
OC   Viruses; Retroid viruses; Retroviridae; Gammaretrovirus.
OX   NCBI_TaxID=11787;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87207680; PubMed=3033897;
RA   Merregaert J., Janowski M., Reddy E.P.;
RT   "Nucleotide sequence of a radiation leukemia virus genome.";
RL   Virology 158:88-102(1987).
RN   [2]
RP   SEQUENCE OF 478-505.
RX   MEDLINE=81264245; PubMed=6267042;
RA   Henderson L.E., Copeland T.D., Sowder R.C., Smythers G.W.,
RA   Oroszlan S.;
RT   "Primary structure of the low molecular weight nucleic acid-binding
RT   proteins of murine leukemia viruses.";
RL   J. Biol. Chem. 256:8400-8406(1981).
CC   -!- SIMILARITY: CONTAINS 1 CCHC-TYPE ZINC FINGER.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; K03363; AAA46518.1; -.
DR   PIR; A26183; FOMVRV.
DR   InterPro; IPR000840; Gag_MA.
DR   InterPro; IPR002079; Gag_p12.
DR   InterPro; IPR003036; Gag_p30.
DR   InterPro; IPR001878; Znf_CCHC.
DR   Pfam; PF01140; gag_MA; 1.
DR   Pfam; PF01141; gag_p12; 1.
DR   Pfam; PF02093; Gag_p30; 1.
DR   Pfam; PF00098; zf-CCHC; 1.
DR   SMART; SM00343; ZnF_C2HC; 1.
DR   PROSITE; PS50158; ZF_CCHC; 1.
KW   Core protein; Coat protein; Nucleoprotein; Polyprotein; Myristate;
KW   Zinc-finger.
FT   CHAIN         2    129       CORE PROTEIN P15.
FT   CHAIN       130    214       INNER COAT PROTEIN P12.
FT   CHAIN       215    477       CORE SHELL PROTEIN P30.
FT   CHAIN       478    537       NUCLEOPROTEIN P10.
FT   ZN_FING     501    518       CCHC-TYPE.
```

```
FT   LIPID         2      2       MYRISTATE.
FT   CONFLICT    479    479       T -> S (IN REF. 2).
FT   CONFLICT    482    482       T -> S (IN REF. 2).
SQ   SEQUENCE   537 AA;  60784 MW;  312AF7B2BBB4B7FB CRC64;


  Query Match             6.3%;  Score 135;  DB 1;  Length 537;
  Best Local Similarity   19.1%;  Pred. No. 0.11;
  Matches   97;  Conservative   58;  Mismatches  173;  Indels  180;  Gaps   20;

Qy      21 PGLHPHRS--------EATAGRSPP------TPTVTLGPDCP-----PPPPPPPPNNNNN 61
           || | |         || |   |       :| ::| |  |     |   |||
Db      71 PGPHGHPDQVPYIVTWEAIAYEPPSWVKPFVSPKLSLSPTAPILPSGPSTQPPP------ 124

Qy      62 NNSKHTGHKSACVPNMTERRRDELSEEINNLREKVMKQSEENNNLQSQVQKLTEENTTLR 121
             :||  | :|           ::: :  |    ::|    :  |||:
Db     125 --------RSALYPALTP----------SIKPRPSKPQVLSDNGGPLIDLLTEDPPPYG 165

Qy     122 EQVEPTPEDEDDDIELRGAAAAAAP-----------PPPIEEECPEDLPEKFDGNPDM- 168
           ||   :|: : |  |    :   ||           ||  :       | :  ||  :
Db     166 EQGPSSPDGDGDREEATYTSEIPAPSPMVSRLRGKRDPPAADSTTSRAFPLRLGGNGQLQ 225

Qy     169 LAPFMAQCQIFMEKSTRDFSVDRVRV-CFVTSMMTGRAARWASAKLERSHYLMHNYPAFM 227
             || :      : :   || | ::   : |::|     |   :                :
Db     226 YWPFSSSDLYNWKNNNPSFSEDPGKLTALIESVLTTHQPTWDDCQQ----------L 272

Qy     228 MEMKHVFEDPQRREVAKRKIRRLRQGMGSVI--DYSNAFQMIAQDLDWNEP--------- 276
           :      |: ||  :  ||  |   |  : :  : ::|| :   | |:  |
Db     273 LGTLLTGEEKQRVLLEARKAVRGNDGRPTQLPNEVNSAFPLERPDWDYTTPEGRNHLVLY 332

Qy     277 --------------------------------ALIDQYHEGL--------SDHIQE--- 292
                                            | :::  |           || ||
Db     333 RQLLLAGLQNAGRSPTNLAKVKGITQGPNESPSAFLERLKEAYRRYTPYDPEDHGQETSV 392

Qy     293 --------------ELSHLE--VAKSLSALIGQCIHI----------------------E 314
                     :|  ||   :|:|  |: :   |                        |
Db     393 SMSFIWQSAPDIGRKLERLEDLKSKTLRDLVREAEKIFNKRETPEEREERFRRETEENEE 452

Qy     315 RRLARAAAAARKPRSPPR----ALVLPHIASHHQVDPTEPVGGARMRLTQEEKERRRKLNL 370
           || |      | |   |    : :|  : :  : |     || | | . | :|::
Db     453 RRRAEDEQREKERDRRRQREMSKLLATVVTGQRQDRQ---GGERKR-PQLDKDQ------ 502

Qy     371 CLYCGTGGHYADNCPAKASKSSPAGNSP 398
           | ||   ||:| :|| |     | |  |
Db     503 CAYCKEKGHWAKDCPKK--PRGPRGPRP 528


RESULT   4
GAG_MLVFF
ID   GAG_MLVFF      STANDARD;      PRT;   538 AA.
AC   P26806;
DT   01-AUG-1992 (Rel. 23, Created)
DT   01-AUG-1992 (Rel. 23, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   GAG polyprotein (Core polyprotein) [Contains: Matrix protein P15; RNA
DE   binding phosphoprotein P12; Capsid protein P30; Nuclecapsid protein
DE   P10].
GN   GAG.
OS   Friend murine leukemia virus (isolate FB29) (F-MuLV).
OC   Viruses; Retroid viruses; Retroviridae; Mammalian type C retroviruses.
OX   NCBI_TaxID=11797;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92107687; PubMed=1762923;
RA   Perryman S., Nishio J., Chesebro B.;
RT   "Complete nucleotide sequence of Friend murine leukemia virus, strain
RT   FB29.";
RL   Nucleic Acids Res. 19:6950-6950(1991).
RN   [2]
RP   SEQUENCE OF 479-512.
RX   MEDLINE=81264245; PubMed=6267042;
RA   Henderson L.E., Copeland T.D., Sowder R.C., Smythers G.W.,
RA   Oroszlan S.;
RT   "Primary structure of the low molecular weight nucleic acid-binding
RT   proteins of murine leukemia viruses.";
RL   J. Biol. Chem. 256:8400-8406(1981).
CC   -!- PTM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC   -!- MISCELLANEOUS: THIS PROTEIN IS SYNTHESIZED AS A GAG-POL
CC       POLYPROTEIN.
CC   -!- SIMILARITY: CONTAINS 1 CCHC-TYPE ZINC FINGER.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z11128; CAA77478.1; -.
DR   InterPro; IPR000840; Gag_MA.
DR   InterPro; IPR002079; Gag_p12.
DR   InterPro; IPR003036; Gag_p30.
DR   InterPro; IPR001878; Znf_CCHC.
DR   Pfam; PF01140; gag_MA; 1.
DR   Pfam; PF01141; gag_p12; 1.
DR   Pfam; PF02093; Gag_p30; 1.
DR   Pfam; PF00098; zf-CCHC; 1.
DR   SMART; SM00343; ZnF_C2HC; 1.
DR   PROSITE; PS50158; ZF_CCHC; 1.
KW   Coat protein; Core protein; Polyprotein; Nucleoprotein; Myristate;
KW   Phosphorylation; Zinc-finger.
FT   CHAIN         2    131       MATRIX PROTEIN P15.
FT   CHAIN       132    215       RNA BINDING PHOSPHOPROTEIN P12.
FT   CHAIN       216    478       CAPSID PROTEIN P30.
FT   CHAIN       479    538       NUCLEOCAPSID PROTEIN P10.
FT   ZN_FING     502    519       CCHC-TYPE.
FT   LIPID         2      2       MYRISTATE (BY SIMILARITY).
FT   CONFLICT    501    501       H -> R (IN REF. 2).
SQ   SEQUENCE   538 AA;  60929 MW;  2E652DDB9E0D4A3C CRC64;


  Query Match             6.1%;  Score 131.5;  DB 1;  Length 538;
  Best Local Similarity   18.8%;  Pred. No. 0.18;
  Matches   95;  Conservative   53;  Mismatches  185;  Indels  171;  Gaps   17;

Qy      21 PGLHPHRS--------EATAGRSP--------PTPTVTLGPDCPPPPPPPPPNNNNNNNS 64
           || | |         || |   |        | | ::| |  |  || || :
Db      71 PGPHGHPDQVPYIVTWEAIAVDPPPWVRPFVHPKPPLSLPPSAPSLPPEPPLS------- 123

Qy      65 KHTGHKSACVPNMTERRRDELSEEINNLREKVMKQSEENNNLQSQVQKLTEENTTLREQV 124
             |  :|:  | :|       |      | :|:  |         :  |||:    |:
Db     124 --TPPQSSLYPALT-------SPLNTKPRPQVLPDSG-----GPLIDLLTEDPPPYRDPG 169

Qy     125 EPTPEDEDDDIELRGAAAAAAP-----------PPPIEEECPEDLPEKFDGNPD-MLAP 171
            |:|:   |  |:     | |            ||  :    :  | :  ||      |
Db     170 PPSPDGNGDSGEVAPTEGAPDPSPMVSRLRGRKEPPVADSTTSQAFPLRLGGNGQYQYWP 229

Qy     172 FMAQCQIFMEKSTRDFSVDRVRV-CFVTSMMTGRAARWASAKLERSHYLMHNYPAFMMEM 230
           | :      : :   || | ::   : |::      |   :                ::
Db     230 FSSSDLYNWKNNNPSFSEDPAKLTALIESVLLTHQPTWDDCQQ------------LLGT 276

Qy     231 KHVFEDPQRREVAKRKIRRLRQGMGSVI--DYSNAFQMIAQDLDWN-------------- 274
             |: ||  :  ||  |   |  : :  | ::|| :   | |:|
Db     277 LLTGEEKQRVLLEARKAVRGEDGRPTQLPNDINDAFPLERPDWDYNTQRGRNHLVHYRQL 336

Qy     275 ----------------------------EPALIDQYHEGL---------------------- 286
                                        | :::  |
Db     337 LLAGLQNAGRSPTNLAKVKGITQGPNESPSAFLERLKEAYRRYTPYDPEDPGQETNVAMS 396

Qy     287 -----SDHIQEELSHLE--VAKSLSALIGQCIHI----------------------ERRL 317
                :  |  :|  ||   :|:|  |: :   |                      |||
Db     397 FIWQSAPDIGRKLERLEDLKSKTLGDLVREAEKIFNKRETPEEREERIRRETEEKEERRR 456

Qy     318 ARAAAARKPRSPPRALVLPHIASHHQVDPTEPVGGARMRLTQEEKERRRKL---NLCLYC 374
           |      | |   |         | ::        :  |  ::  ||||     : | ||
```

```
Db    457 AEDVQREKERDRRR---------HREMSKLLATVVSGQRQDRQGGERRRPQLDHDQCAYC 507

Qy    375 GTGGHYADNCPAKASKSSPAGNSP 398
          ||:| :|| |     | |  |
Db    508 KEKGHWARDCPKK--PRGPRGPRP 529


RESULT  5
ST20_CANAL
ID   ST20_CANAL     STANDARD;      PRT;  1230 AA.
AC   Q92212;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Serine/threonine-protein kinase STE20 homolog (EC 2.7.1.-).
GN   HST20 OR CST20.
OS   Candida albicans (Yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX   NCBI_TaxID=5476;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1066;
RX   MEDLINE=97075146; PubMed=8917572;
RA   Kohler J.R., Fink G.R.;
RT   "Candida albicans strains heterozygous and homozygous for mutations
RT   in mitogen-activated protein kinase signaling components have defects
RT   in hyphal development.";
RL   Proc. Natl. Acad. Sci. U.S.A. 93:13223-13228(1996).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC       STE20 SUBFAMILY.
CC   -!- SIMILARITY: CONTAINS 1 GBD DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U73457; AAB38875.1; -.
DR   HSSP; P00518; 1PHK.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR000095; PAK_box_P21_Rho_bindng.
DR   InterPro; IPR002290; Ser_thr_pkinase.
DR   Pfam; PF00786; PBD; 1.
DR   Pfam; PF00069; pkinase; 1.
DR   SMART; SM00285; PBD; 1.
DR   SMART; SM00220; S_TKc; 1.
DR   PROSITE; PS50108; GBD; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
KW   Transferase; Serine/threonine-protein kinase; ATP-binding.
FT   DOMAIN      109    119       POLY-ASN.
FT   DOMAIN      290    293       POLY-PRO.
FT   DOMAIN      453    464       POLY-SER.
FT   DOMAIN      552    555       POLY-SER.
FT   DOMAIN      573    580       POLY-GLY.
FT   DOMAIN      727    732       POLY-PRO.
FT   DOMAIN      790    794       POLY-PRO.
FT   DOMAIN      475    532       GBD.
FT   DOMAIN      953   1205       PROTEIN KINASE.
FT   NP_BIND     959    967       ATP (BY SIMILARITY).
FT   BINDING     983    983       ATP (BY SIMILARITY).
FT   ACT_SITE   1073   1073       BY SIMILARITY.
SQ   SEQUENCE   1230 AA;  132862 MW;  2B2AC4C133B9FE81 CRC64;


  Query Match             6.1%;  Score 131.5;  DB 1;  Length 1230;
  Best Local Similarity  21.5%;  Pred. No. 0.46;

  Matches   64;  Conservative   51;  Mismatches  121;  Indels   61;  Gaps   13;

Qy     17 GGQDPGLHPHR-SEATAGRSPPTPTVTLGPDCPPPPPPPPPNNNNNNNSKHTGHKSA--- 72
          ||::  |   | :|  | |:|| |     |  || || || |  :   |:    ::|
Db    770 GGENNALPKQRINEFKAHRAPPPP--PSAPPAPPVPPAPPANLLSEQTSEIPQQRTAPSQ 827

Qy     73 ------CVPNMTERRRDELSEEINNLREKVMKQSEENNNLQSQVQKLTEENTTLREQVEP 126
                  |: | :: :  |    ||||  :: ||       :|:| |:|     | |
Db    828 ALADVTAPTNIYEIQQTKYQEAQQKLREKKARELEE-------IQRLREKNERQNRQQET 880

Qy    127 TPEDEDDDIELRGAAAAAAPPPPIEEECPEDLPEKFDGNPD-MLAPFMAQ---------- 175
            : |         :  ||| |:  : |   |    |  |   |  :||
Db    881 GQNNAD----TASGGSNIAPPVPVPNKKP---PSGSGGGRDAKQAALIAQKKREEKKRKN 933

Qy    176 CQIFMEKST-------RDFSVDRVRVCFVTSMMTGRAARWASAKLERSHYLMHNYPAFMM 228
           ||  : |         :  || |::            : ||  :  :| :        :
Db    934 LQIIAKLKTICNPGDPNELYVDLVKI----------GQGASGGVFLAHDVRDKSNIVAI 982

Qy    229 EMKHVFEDPQRREVAKRKIRRLRQGMGSVIDYSNAFQMIAQDLDWNEPALIDQYHEG 285
          :  :: : |::  :    :        ::::: ::: ::  || |    :| :| ||
Db    983 KQMNLEQQPKKELIINEILVMKGSSHPNIVNFIDSY-LLKGDL-W----VIMEYMEG 1033


RESULT  6
GAG_MLVBM
ID   GAG_MLVBM      STANDARD;      PRT;   537 AA.
AC   P29167;
DT   01-DEC-1992 (Rel. 24, Created)
DT   01-DEC-1992 (Rel. 24, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   GAG polyprotein [Contains: Core protein P15; Inner coat protein P12;
DE   Core shell protein P30; Nucleoprotein P10].
GN   GAG.
OS   Murine leukemia virus (strain BM5 eco).
OC   Viruses; Retroid viruses; Retroviridae; Gammaretrovirus.
OX   NCBI_TaxID=31687;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91303677; PubMed=1649328;
RA   Chattopadhyay S.K., Sengupta D.N., Fredrickson T.N., Morse H.C. III,
RA   Hartley J.W.;
RT   "Characteristics and contributions of defective, ecotropic, and mink
RT   cell focus-inducing viruses involved in a retrovirus-induced
RT   immunodeficiency syndrome of mice.";
RL   J. Virol. 65:4232-4241(1991).
CC   -!- PTM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC   -!- SIMILARITY: CONTAINS 1 CCHC-TYPE ZINC FINGER.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M64095; AAA46510.1; -.
DR   PIR; A40416; FOMVMB.
DR   InterPro; IPR000840; Gag_MA.
DR   InterPro; IPR002079; Gag_p12.
DR   InterPro; IPR003036; Gag_p30.
DR   InterPro; IPR001878; Znf_CCHC.
DR   Pfam; PF01140; gag_MA; 1.
DR   Pfam; PF01141; gag_p12; 1.
DR   Pfam; PF02093; Gag_p30; 1.
DR   Pfam; PF00098; zf-CCHC; 1.
DR   SMART; SM00343; ZnF_C2HC; 1.
DR   PROSITE; PS50158; ZF_CCHC; 1.
KW   Coat protein; Core protein; Polyprotein; Nucleoprotein; Myristate;
KW   Zinc-finger.
FT   CHAIN         2    129       CORE PROTEIN P15.
FT   CHAIN       130    214       INNER COAT PROTEIN P12.
FT   CHAIN       215    477       CORE SHELL PROTEIN P30.
```

```
FT   CHAIN       478    537       NUCLEOPROTEIN P10.
FT   ZN_FING     501    518       CCHC-TYPE.
FT   LIPID         2      2       MYRISTATE (BY SIMILARITY).
SQ   SEQUENCE   537 AA;  60422 MW;  ABD2E70299BFFD64 CRC64;


  Query Match             6.1%;  Score 130;  DB 1;  Length 537;
  Best Local Similarity   18.5%;  Pred. No. 0.22;
  Matches   93;  Conservative   53;  Mismatches  186;  Indels  172;  Gaps   17;

Qy      21 PGLHPHRS--------EATAGRSPP------TPTVTLGPDCP-----PPPPPPPPNNNNN 61
           || | |          || |   ||      :| ::| |  |     |   |||
Db      71 PGPHGHPDQVPYIVTWEAIAYEPPPWVKPFVSPKLSLSPTAPILPSGPSTQPPP------ 124

Qy      62 NNSKHTGHKSACVPNMTERRRDELSEEINNLREKVMKQSEENNNLQSQVQKLTEENTTLR 121
              :||  |  |          |     |    |::   |   :  |||:
Db     125 --------RSALYPAFT--------PSIKPRPSKPQVLSDDGGPL---IDLLTEDPPPYG 165

Qy     122 EQVEPTPEDEDDDIELRGAAAAAAP------------PPPIEEECPEDLPEKFDGNPDM- 168
           ||   :|: : |  |    :   ||             ||  :       | :  ||  :
Db     166 EQGPSSPDGDGDREEATSTSEIPAPSPMVSRLRGKRDPPAADSTTSRAFPLRLGGNGQLQ 225

Qy     169 LAPFMAQCQIFMEKSTRDFSVDRVRV-CFVTSMMTGRAARWASAKLERSHYLMHNYPAFM 227
             || :      : :   || |  ::   : |::|     |   :              :
Db     226 YWPFSSSDLYNWKNNNPSFSEDPGKLTALIESVLTTHQPTWDDCQQ------------L 272

Qy     228 MEMKHVFEDPQRREVAKRKIRRLRQGMGSVI--DYSNAFQMIAQDLDWNEP--------- 276
           :      |: ||  :  ||  |   |  : :  : ::|| :   | |:  |
Db     273 LGTLLTGEEKQRVLLEARKAVRGNDGRPTQLPNEVNSAFPLERPDWDYTTPEGRNHLVLY 332

Qy     277 --------------------------------ALIDQYHEGLSDH-------------- 289
                                            | :::  |    :
Db     333 RQLLLAGLQNAGRSPTNLAKVKGITQGPNESPSAFLERLKEAYRRYTPYDPEDPGQETNV 392

Qy     290 -----------IQEELSHLE--VAKSLSALIGQCIHI----------------------E 314
                     |  :|  ||   :|:|  |: :   |                       |
Db     393 SMSFIWQSAPAIGRKLERLEDLKSKTLGDLVREAEKIFNKRETPEEREERIRRETEEKEE 452

Qy     315 RRLARAAAARKPRSPPRALVLPHIASHHQVDPTEPVGGARMRLTQEEKERRRKLNLCLYC 374
           || |      | |   |   :  : :       :   |   |  | :|::      | ||
Db     453 RRRAGDEQREKERDRRRQREMSKLLATVVTGQRQDRQGGERRRPQLDKDQ------CAYC 506

Qy     375 GTGGHYADNCPAKASKSSPAGNSP 398
             ||:| :|| |     | |  |
Db     507 KEKGHWAKDCPKK--PRGPRGPRP 528


RESULT   7
S3B2_HUMAN
ID   S3B2_HUMAN     STANDARD;      PRT;   872 AA.
AC   Q13435;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Splicing factor 3B subunit 2 (Spliceosome associated protein 145) (SAP
DE   145) (SF3b150) (Pre-mRNA splicing factor SF3b 145 kDa subunit).
GN   SF3B2 OR SAP145.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 151-159 AND 794-817.
RX   MEDLINE=96154048; PubMed=8566756;
RA   Gozani O., Feld R., Reed R.;
RT   "Evidence that sequence-independent binding of highly conserved U2
RT   snRNP proteins upstream of the branch site is required for assembly
RT   of spliceosomal complex A.";
RL   Genes Dev. 10:233-243(1996).
RN   [2]
RP   CHARACTERIZATION OF THE SPLICEOSOME.
RX   MEDLINE=20337962; PubMed=10882114;
RA   Das R., Zhou Z., Reed R.;
RT   "Functional association of U2 snRNP with the ATP-independent
RT   spliceosomal complex E.";
RL   Mol. Cell 5:779-787(2000).
CC   -!- FUNCTION: SUBUNIT OF THE SPLICING FACTOR SF3B REQUIRED FOR 'A'
CC       COMPLEX ASSEMBLY FORMED BY THE STABLE BINDING OF U2 SNRNP TO THE
CC       BRANCHPOINT SEQUENCE (BPS) IN PRE-MRNA. SEQUENCE INDEPENDENT
CC       BINDING OF SF3A/SF3B COMPLEX UPSTREAM OF THE BRANCH SITE IS
CC       ESSENTIAL, IT MAY ANCHOR U2 SNRNP TO THE PRE-MRNA. MAY ALSO BE
CC       INVOLVED IN THE ASSEMBLY OF THE 'E' COMPLEX. BELONGS ALSO TO THE
CC       MINOR U12-DEPENDENT SPLICEOSOME, WHICH IS INVOLVED IN THE SPLICING
CC       OF RARE CLASS OF NUCLEAR PRE-MRNA INTRON.
CC   -!- SUBUNIT: COMPONENT OF SPLICING FACTOR SF3B WHICH IS COMPOSED OF
CC       FOUR SUBUNITS; SF3B4/SAP49, SF3B3/SAP130, SF3B2/SAP145,
CC       SF3B1/SAP155. SF3B ASSOCIATES WITH THE SPLICING FACTOR SF3A AND A
CC       12S RNA UNIT TO FORM THE U2 SMALL NUCLEAR RIBONUCLEOPROTEINS
CC       COMPLEX (U2 SNRNP). SF3B2 INTERACTS DIRECTLY WITH SF3B4.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR (By similarity).
CC   -!- SIMILARITY: TO YEAST CUS1. SOME, TO C.ELEGANS ZK632.11.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; U41371; AAA97461.1; -.
DR   MIM; 605591; -.
DR   InterPro; IPR003034; SAP.
DR   Pfam; PF02037; SAP; 1.
DR   SMART; SM00513; SAP; 1.
KW   Spliceosome; mRNA processing; mRNA splicing; Nuclear protein.
FT   DOMAIN       68     73       POLY-PRO.
FT   DOMAIN       81     90       POLY-PRO.
FT   DOMAIN      106    109       POLY-PRO.
FT   DOMAIN      226    230       POLY-PRO.
FT   DOMAIN      269    274       POLY-GLU.
FT   DOMAIN      308    312       POLY-LYS.
FT   DOMAIN      676    679       POLY-GLU.
FT   DOMAIN      697    703       POLY-GLU.
SQ   SEQUENCE   872 AA;  97656 MW;  AED669FD0DA5DE31 CRC64;


  Query Match             6.0%;  Score 129;  DB 1;  Length 872;
  Best Local Similarity   20.7%;  Pred. No. 0.45;
  Matches   83;  Conservative   64;  Mismatches  150;  Indels  104;  Gaps   19;

Qy      16 RGGQDPG-------LHPHRSEATAGRSPPTPTV-TLGPDCPPP--PPPPPPNNNNNNNSK 65
           |  || |       | | |  |  |  |||||  :|   | |  ||||| : |   :
Db     182 RPPQDMGQIGVRTPLGP-RVAAPVGPVGPTPTVLPMGAPVPRPRGPPPPPGDENREMDDP 240

Qy      66 HTGHKSACVPNMTER---RRDELSEEINNLREKVMKQSEENNNLQSQVQKLTEENTTLRE 122
             | |   :|   |:    ::   ||:|: :|:   :::  ::|     :  |:  :: :
Db     241 SVGPK---IPQALEKILQLKESRQEEMNSQQEEEEMETDARSSLGQSASETEEDTVSVSK 297

Qy     123 Q----------------------VEPTPEDEDDDIELRGAAAAAAPPPPIEEECPEDLPE 160
           :                      | : : | |    ||: : ||    :| |   : ||
Db     298 KEKNRKRRNRKKKKKPQRVRGVSSESSGDREKDSTRSRGSDSPAA---DVEIEYVTEEPE 354

Qy     161 KFDGN----PDMLAPFMAQCQIFMEKSTRDFSVDRVRVCFVTSMMTGRAARWASAKLERS 216
           :: |      :   |   :  ||     :|                          ||| |
Db     355 IYEPNFIFFKRIFEAFKLTDDVKKEKEKEPEKLD--------------------KLENS 393

Qy     217 HYLMHNYPAFMMEMKHVFED---------PQRREVAKRKIRRLRQGMGSVIDYSNAFQMI 267
                 |  | |   :|          |:  :::|:|:||: :         :   |::
Db     394 --AAPKKKGFEEEHKDSDDDSSDDEQEKKPEAPKLSKKKLRRMNR-----FTVAELKQLV 446

Qy     268 AQDLDWNEPALIDQYHEGLSDHIQEELSHLEVAKSLSALIGQCIHIERRLARAAAARKPR 327
           |:      | ::: :     |   : | ||:  :: |  : :    :|:  :    |
Db     447 AR------PDVVEMHDVTAQD--PKLLVHLKATRN-SVPVPRHWCFKRKYLQ--GKRGIE 495

Qy     328 SPPRALVLPHIASHHQVDPTEPVGGARMRLTQEEKERRRKL 368
```

```
          ||    ||       |  :  |   ||   :||| :: :
Db    496 KPP--FELP--------DFIKRTGIQEMREALQEKEEQKTM 526


RESULT   8
GAG_MLVDE
ID   GAG_MLVDE      STANDARD;      PRT;   536 AA.
AC   P29168;
DT   01-DEC-1992 (Rel. 24, Created)
DT   01-DEC-1992 (Rel. 24, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   GAG polyprotein [Contains: Core protein P15; Inner coat protein P12;
DE   Core shell protein P30; Nucleoprotein P10].
GN   GAG.
OS   Murine leukemia virus (strain DEF27).
OC   Viruses; Retroid viruses; Retroviridae; Gammaretrovirus.
OX   NCBI_TaxID=31688;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91303677; PubMed=1649328;
RA   Chattopadhyay S.K., Sengupta D.N., Fredrickson T.N., Morse H.C. III,
RA   Hartley J.W.;
RT   "Characteristics and contributions of defective, ecotropic, and mink
RT   cell focus-inducing viruses involved in a retrovirus-induced
RT   immunodeficiency syndrome of mice.";
RL   J. Virol. 65:4232-4241(1991).
CC   -!- PTM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC   -!- SIMILARITY: CONTAINS 1 CCHC-TYPE ZINC FINGER.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M64096; AAA46509.1; ALT_SEQ.
DR   PIR; B40416; FOMVME.
DR   InterPro; IPR000840; Gag_MA.
DR   InterPro; IPR002079; Gag_p12.
DR   InterPro; IPR003036; Gag_p30.
DR   InterPro; IPR001878; Znf_CCHC.
DR   Pfam; PF01140; gag_MA; 1.
DR   Pfam; PF01141; gag_p12; 1.
DR   Pfam; PF02093; Gag_p30; 1.
DR   Pfam; PF00098; zf-CCHC; 1.
DR   SMART; SM00343; ZnF_C2HC; 1.
DR   PROSITE; PS50158; ZF_CCHC; 1.
KW   Coat protein; Core protein; Polyprotein; Nucleoprotein; Myristate;
KW   Zinc-finger.
FT   CHAIN         2    129       CORE PROTEIN P15.
FT   CHAIN       130    213       INNER COAT PROTEIN P12.
FT   CHAIN       214    476       CORE SHELL PROTEIN P30.
FT   CHAIN       477    536       NUCLEOPROTEIN P10.
FT   ZN_FING     500    517       CCHC-TYPE.
FT   LIPID         2      2       MYRISTATE (BY SIMILARITY).
SQ   SEQUENCE   536 AA;  60890 MW;  4D990FD5C5AEF39B CRC64;


  Query Match             5.7%;  Score 122.5;  DB 1;  Length 536;
  Best Local Similarity   18.8%;  Pred. No. 0.65;
  Matches   97;  Conservative   45;  Mismatches  178;  Indels  195;  Gaps   17;

Qy      21 PGLHPHRS--------EATAGRSP----------PTPTVTLGPDCPPPPPPPPPNNNNNN 62
           || | |          || |   |          | |  || |  || |   ||:
Db      71 PGPHGHPDQVPYIVTWEALAYHPPPWVKPFVSPKPFPLSTL-PFSPPGPSAHPPS----- 124

Qy      63 NSKHTGHKSACVPNMTERRRDELSEEINNLREKVMKQSEENNNLQSQVQKLTEENTTLRE 122
                        | |  :  | ::: |  |       |     :  |||    | |
Db     125 -------------RSDLYTALIPSIKTKPPKSRVLPTNGGPLIDLLTENPPNLGE 166

Qy     123 QVEPTPEDEDDDIELRGAAAAAAPPPPIEEECPEDLPEKFDGNPDMLA------------ 170
```

```
          |  | |         :|      ||||     |  :  :  || |  |
Db    167 QGPPLP---------KGPVKKRRPPPP-RYSPPGPMVSRLRGNRDPPAADSTTSRAFPLR 216

Qy    171 ----------PFMAQCQIFMEKSTRDFSVDRVRV-CFVTSMMTGRAARWASAKLERSHYL 219
                   || :      : :   || |  ::   : |::|     |   :
Db    217 LGGNGQLQYWPFSSSDLYNWKNNNPSFSEDPGKLTALIESVLTTHQPTWDDCQQ------ 270

Qy    220 MHNYPAFMMEMKHVFEDPQRREVAKRKIRRLRQGMGSVI--DYSNAFQMIAQDLDWNEP- 276
                ::      |: ||  :  ||  |   |  : :  : ::|| :   | |:  |
Db    271 -------LLGTLLTGEEKQRVLLEARKAVRGNDGRPTQLPNEVNSAFPLERPDWDYTTPE 323

Qy    277 ----------------------------------------ALIDQYHEGL--------- 286
                                                     | :::  |
Db    324 GRNHLVLYRQLLLAGLQNAGRSPTNLAKVKGITQGPSESPSAFLERLKEAYRRYTPYDPE 383

Qy    287 ----------------SDHIQEELSHLE--VAKSLSALIGQCIHI---------------- 313
                           :  |  :|  ||   :|:|  |: :   |
Db    384 DPGQETNVSMSFIWQSAPDIGRKLERLEDLKSKTLGDLVREAEKIFNKRETPEEREERIR 443

Qy    314 -------ERRLARAAAARKPRSPPRALVLPHIASHHQVDPTEPVGGARMRLTQEEKERRR 366
                  |||  |       | |   |          | ::           |  ::  ||||
Db    444 RETEEKEERRRAEDEQREKERDRRR---------HREMSKFLATVVTGQRQDRQGGERRR 494

Qy    367 ---KLNLCLYCGTGGHYADNCPAKASKSSPAGNSP 398
               : | ||   ||:| :|| |     | | |
Db    495 PQLDKDQCAYCKEKGHWAKDCPKK--PRGPRGPRP 527


RESULT   9
GAG_MLVFP
ID   GAG_MLVFP      STANDARD;      PRT;   538 AA.
AC   P26805;
DT   01-AUG-1992 (Rel. 23, Created)
DT   01-AUG-1992 (Rel. 23, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   GAG polyprotein (Core polyprotein) [Contains: Matrix protein P15; RNA
DE   binding phosphoprotein P12; Capsid protein P30; Nuclecapsid protein
DE   P10].
GN   GAG.
OS   Friend murine leukemia virus (isolate PVC-211) (F-MuLV).
OC   Viruses; Retroid viruses; Retroviridae; Mammalian type C retroviruses.
OX   NCBI_TaxID=11798;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92319660; PubMed=1620621;
RA   Remington M.P., Hoffman P.M., Ruscetti S.K., Masuda M.;
RT   "Complete nucleotide sequence of a neuropathogenic variant of Friend
RT   murine leukemia virus PVC-211.";
RL   Nucleic Acids Res. 20:3249-3249(1992).
CC   -!- PTM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC   -!- MISCELLANEOUS: THIS PROTEIN IS SYNTHESIZED AS A GAG-POL
CC       POLYPROTEIN.
CC   -!- SIMILARITY: CONTAINS 1 CCHC-TYPE ZINC FINGER.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M93134; AAA46476.1; -.
DR   PIR; S35474; S35474.
DR   InterPro; IPR000840; Gag_MA.
DR   InterPro; IPR002079; Gag_p12.
DR   InterPro; IPR003036; Gag_p30.
DR   InterPro; IPR001878; Znf_CCHC.
DR   Pfam; PF01140; gag_MA; 1.
DR   Pfam; PF01141; gag_p12; 1.
DR   Pfam; PF02093; Gag_p30; 1.
DR   Pfam; PF00098; zf-CCHC; 1.
DR   SMART; SM00343; ZnF_C2HC; 1.
```

```
DR   PROSITE; PS50158; ZF_CCHC; 1.
KW   Coat protein; Core protein; Polyprotein; Nucleoprotein; Myristate;
KW   Phosphorylation; Zinc-finger.
FT   CHAIN         2    131       MATRIX PROTEIN P15.
FT   CHAIN       132    215       RNA BINDING PHOSPHOPROTEIN P12.
FT   CHAIN       216    478       CAPSID PROTEIN P30.
FT   CHAIN       479    538       NUCLEOCAPSID PROTEIN P10.
FT   ZN_FING     502    519       CCHC-TYPE.
FT   LIPID         2      2       MYRISTATE (BY SIMILARITY).
SQ   SEQUENCE   538 AA;  61033 MW;  2FF9F97D2C79DEBE CRC64;


  Query Match             5.7%;  Score 122.5;  DB 1;  Length 538;
  Best Local Similarity   18.6%;  Pred. No. 0.65;
  Matches   98;  Conservative   49;  Mismatches  164;  Indels  215;  Gaps   20;

Qy      21 PGLHPHRS--------EATAGRSP--------PTPTVTLGPDCPPPPPPPPPNNNNNNNS 64
           || | |          || |   |         | | : | |  |  || || :
Db      71 PGPHGHPDQVPYIVTWEALAVDPPPWVKPFVHPKPPLLLPPSAPSLPPEPPLS------- 123

Qy      65 KHTGHKSACVPNMTERRRDELSEEINNLREKVMKQSEENNNLQSQVQKLTEENTTLREQV 124
             |  :|:  | :|       |      | :|:  |          :  |||:    |:
Db     124 --TPPQSSLYPALT-------SPLNTKPRPQVLPDSG-----GPLIDLLTEDPPPYRDPG 169

Qy     125 EPTPEDEDDDIELRGAAAAAAP--------------PPPIEEECPEDLPEKFDGNPD-ML 169
            |:|: : |  |:  |    ||              ||  :    :  | :  ||
Db     170 PPSPDGKGDSGEV--APTEGAPDSSPMVSRLRGRREPPVADSTTSQAFPLRLGGNGQFQY 227

Qy     170 APFMAQCQIFMEKSTRDFSVDRVRV-CFVTSMMTGRAARWASAKLERSHYLMHNYPAFMM 228
            || :      : :   || | ::   : |::      |   :              ::
Db     228 WPFSSSDLYNWKNNNPSFSEDPGKLTALIESVLLTHQPTWDDCQQ------------LL 274

Qy     229 EMKHVFEDPQRREVAKRKIRRLRQGMGSVI--DYSNAFQMIAQDLDWNEPALIDQYHEGL 286
               |: ||  :  ||  |   |  : :  | ::|| :   | |:|
Db     275 GTLLTGEEKQRVLLEARKAVRGEDGRPTQLPNDINDAFPLERPDWDYN------------ 322

Qy     287 SDHIQEELSHLEVAKSLSALIGQCIHIERRLARAAAARKPRSPPRALVLPHIAS------ 340
             |   :||             :|  |:|  |      |||     :  |
Db     323 ---TQRGRNHL------------VHY-RQLLLAGLQNAGRSPTNLAKVKGITQGPNESP 365

Qy     341 -------------HHQVDPTEP----------------VGGARMRL------------- 357
                       :  || :|              :|    ||
Db     366 SAFLERLKEAYRRYTPYDPEDPGQETNVSMSFIWQSAPDIGRKLERLEDLKNKTLGDLVR 425

Qy     358 ----------------------TQEEKERRR------------------KL-------- 368
                                  |:|::||||                  ||
Db     426 EAEKIFNKRETPEEREERVRRETEEKEERRRAEDERREKERDRRRHREMSKLLATVVSGQ 485

Qy     369 ---------------NLCLYCGTGGHYADNCPAKASKSSPAGNSP 398
                          : | ||   ||:| :|| |     | |  |
Db     486 RQDRQGGERRRPQLDHDQCAYCKEKGHWARDCPKK--PRGPRGPRP 529


RESULT  10
P3K3_DICDI
ID   P3K3_DICDI     STANDARD;      PRT;  1585 AA.
AC   P54675;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Phosphatidylinositol 3-kinase 3 (EC 2.7.1.137) (PI3-kinase)
DE   (PtdIns-3-kinase) (PI3K) (Fragment).
GN   PIKC OR PIK3.
OS   Dictyostelium discoideum (Slime mold).
OC   Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX   NCBI_TaxID=44689;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AX3;
RX   MEDLINE=96009592; PubMed=7565716;
RA   Zhou K., Takegawa K., Emr S.D., Firtel R.A.;
RT   "A phosphatidylinositol (PI) kinase gene family in Dictyostelium
RT   discoideum: biological roles of putative mammalian p110 and yeast
RT   Vps34p PI 3-kinase homologs during growth and development.";
RL   Mol. Cell. Biol. 15:5645-5656(1995).
CC   -!- CATALYTIC ACTIVITY: ATP + 1-phosphatidyl-1D-myo-inositol = ADP +
CC       1-phosphatidyl-1D-myo-inositol 3-phosphate.
CC   -!- SIMILARITY: BELONGS TO THE PI3/PI4-KINASES FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U23478; AAA85723.1; -.
DR   DictyDb; DD01101; pikC.
DR   InterPro; IPR002420; PI3K_C2.
DR   InterPro; IPR000341; PI3K_ras_bind.
DR   InterPro; IPR001263; PI3Ka.
DR   InterPro; IPR000403; PI3_PI4_kinase.
DR   Pfam; PF00613; PI3Ka; 1.
DR   Pfam; PF00792; PI3K_C2; 1.
DR   Pfam; PF00794; PI3K_rbd; 1.
DR   Pfam; PF00454; PI3_PI4_kinase; 1.
DR   SMART; SM00142; PI3K_C2; 1.
DR   SMART; SM00144; PI3K_rbd; 1.
DR   SMART; SM00145; PI3Ka; 1.
DR   SMART; SM00146; PI3Kc; 1.
DR   PROSITE; PS00915; PI3_4_KINASE_1; 1.
DR   PROSITE; PS00916; PI3_4_KINASE_2; 1.
DR   PROSITE; PS50290; PI3_4_KINASE_3; 1.
KW   Transferase; Kinase; Multigene family; Repeat.
FT   NON_TER       1      1
FT   DOMAIN       58     84       POLY-ASN.
FT   DOMAIN       98    110       POLY-ASN.
FT   DOMAIN      200    226       POLY-ASN.
FT   DOMAIN      239    254       POLY-ASN.
FT   DOMAIN      345    378       POLY-ASN.
FT   DOMAIN      383    390       POLY-ASN.
FT   DOMAIN      484    488       POLY-SER.
FT   DOMAIN      720    737       POLY-GLN.
FT   DOMAIN     1569   1575       POLY-ASN.
FT   DOMAIN     1221   1484       PI3K/PI4K.
FT   DOMAIN     1510   1539       5 X 5 AA APPROXIMATE REPEATS.
FT   REPEAT     1510   1514       1.
FT   REPEAT     1515   1519       2.
FT   REPEAT     1520   1524       3.
FT   REPEAT     1530   1534       4.
FT   REPEAT     1535   1539       5.
FT   DOMAIN     1547   1560       7 X 2 AA TANDEM REPEATS OF K-E.
SQ   SEQUENCE   1585 AA;  180421 MW;  4689B620D2484961 CRC64;


  Query Match             5.7%;  Score 121.5;  DB 1;  Length 1585;
  Best Local Similarity   21.0%;  Pred. No. 2.7;
  Matches   71;  Conservative   50;  Mismatches  114;  Indels  103;  Gaps   17;

Qy      51 PPPPPP--------NNNNNNNSKHTGHKSACVPNMTERRRDELSEEINNLREKVMKQSEE 102
           |   ||        |||||||: :  :      |      :  :   ||  |::: :
Db     333 PSSSPPTQSDIFNENNNNNNNNNNNNNNN-----NNNNNNNNNNNNNNNNNNEELINNNNN 387

Qy     103 NNNLQSQVQKLTEENTTLREQVEP-----------TPEDEDDDIELRG--------AAA 142
           |||   :  |: |   :|:| :|               :| |::: :          |
Db     388 NNN--DENYKIEETEESLKELLEKEKLENEEREKILKERNEIDNLKKKNHLSKGYFMRAC 445

Qy     143 AAAPPPPIEEECPEDLP---EKFDGNPDMLAPFMAQCQIFMEKSTRDFSVDRVRVCFVTS 199
             |:    :||   ||:|   | :: |  :| |      ::    |:   |:| :
Db     446 NASNDDGLEE---EDIPLQDEHWETNVIVLLPCRHHVKVPGSSSS---SIDSI------- 492

Qy     200 MMTGRAARWASAKLERSHYLMHNYPAFMMEM--KHVFEDPQRREVAKRKIRRLRQGMGSV 257
               |  ||| |::    |  :   | :    | |  |                 :| :
Db     493 ----RQLAWASGKMQGHLNLEKDEKFFTLRWCNKDVVFDQD------------TPLGHL 535
```

```
Qy    258 IDYSNAFQMIAQDLDWNEPA---------LIDQYHEGLSDHIQEE----LSHLEVAKSLS 304
          | |         :|::| |           |:      | | : :|    |  ||:
Db    536 IQY---------NLNYNNPTQKPTNIKLELV------LEDELCKERLVDLQSLEINNGRP 580

Qy    305 AL----IGQCIHIERRLARAAAAARKPRS-PPRALVLPH 337
          ::    |   :   |:|   |   ||:|  | | : |:
Db    581 SIWKSHIDDVLSFNRKLRELAMLAKPQSNVPAARLTPY 618


RESULT 11
Y182_HUMAN
ID   Y182_HUMAN     STANDARD;      PRT;  1157 AA.
AC   Q14687;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical protein KIAA0182 (Fragment).
GN   KIAA0182.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Bone marrow;
RX   MEDLINE=96281124; PubMed=8724849;
RA   Nagase T., Seki N., Ishikawa K.-I., Tanaka A., Nomura N.;
RT   "Prediction of the coding sequences of unidentified human genes. V.
RT   The coding sequences of 40 new genes (KIAA0161-KIAA0200) deduced by
RT   analysis of cDNA clones from human cell line KG-1.";
RL   DNA Res. 3:17-24(1996).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D80004; BAA11499.1; -.
KW   Hypothetical protein.
FT   NON_TER       1      1
FT   DOMAIN      591    596       POLY-PRO.
FT   DOMAIN      685    688       POLY-ARG.
FT   DOMAIN     1042   1047       POLY-GLU.
SQ   SEQUENCE   1157 AA;  130323 MW;  B0DADAE5A6FAB5FB CRC64;


  Query Match             5.6%;  Score 120.5;  DB 1;  Length 1157;
  Best Local Similarity   21.9%;  Pred. No. 2.1;
  Matches   95;  Conservative   52;  Mismatches  124;  Indels  163;  Gaps   25;

Qy     23 LHPHRSEATAGRSPPTPTVT---------LGPDCPPP------PPPPP----PNNNNNNN 63
          ||  |  ||  |  |:  :|             | |      | | |    |:  :|:
Db    354 LHGLRGHATEERGKPSEQLTPTRAEKLKDASLQAPKPVQHPLHPVPTPHHTVPSLISNHG 413

Qy     64 SKHTGHKSACVPNMTERRRDELSEEINNLREKVMKQSEENNNLQSQVQKLTEENTTLREQ 123
             ||    : :|  :   ||    |:: ::| :|:   |||| :  ::   | :
Db    414 IFSLPSSSAATALLIQRTNE---EEKWLARQRRLRQEKEDR--QSQVSEFRQQ--VLEQH 466

Qy    124 VE----PTPEDEDDDIELRGAAAAAAPPPPIEEECPEDLPEKFDGNPDMLAP----FMAQ 175
          ::    | | : :   |     :   | |   |    | |: | | | :::|      |
Db    467 LDMGRPPVPAEAEHRPE-----STTRPGPNRHEPGGRDPPQHFGGPPPLISPKPQLHAAP 521

Qy    176 CQIFMEKSTRDFSVDRVRVCFVTSMMTGRAARWASAKLERSHYLMHNYPAF--------- 226
          ::   |  |           :: | ||          || | |::||
Db    522 TALWNPVSLMD----------NTLETRRA---------ESHSL-HSHPAAFEPSRQAAV 560

Qy    227 -MMEMKHVFEDPQRREVAKRKIRRLRQGMGSVIDYSNAFQMIAQDLDWNEPALIDQY--- 282
           ::::: ||  |:: |   ||                          ||| :|:|
Db    561 PLVKVERVF-CPEKAEEGPRK-------------------------REPAPLDKYQPP 592

Qy    283 ----HEGLSDHIQEELSH----------LEVAKSLSALIGQCIHIERRLARAAAAR--- 324
              || |      | |            |: ||   ::||     :| :   ||
Db    593 PPPPREGGS------LEHQPFLPGPGPFLAELEKSTQTILGQ-----QRASLPQAATFGE 641

Qy    325 -----KPRSP-----PRALVLPHIASHHQVDPT---EPVGGARMRLTQE---EKERRRKL 368
              || ||     |||           ||    :     | ||  :   |: |||:
Db    642 LSGPLKPGSPYRPPVPRA----------PDPAYIYDEFLQQRRRLVSKLDLEERRRREA 690

Qy    369 NLCLYCGTGGHYAD 382
                |:| |
Db    691 Q------EKGYYYD 698 .


RESULT 12
AF6_HUMAN
ID   AF6_HUMAN      STANDARD;      PRT;  1816 AA.
AC   P55196; O75087; O75088; O75089; Q9NU92;
DT   01-OCT-1996 (Rel. 34, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   AF-6 protein.
GN   MLLT4 OR AF6.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RX   MEDLINE=94061833; PubMed=8242616;
RA   Prasad R., Gu Y., Alder H., Nakamura T., Canaani O., Saito H.,
RA   Huebner K., Gale R.P., Nowell P.C., Kuriyama K., Miyazaki Y.,
RA   Croce C.M., Canaani E.;
RT   "Cloning of the ALL-1 fusion partner, the AF-6 gene, involved in
RT   acute myeloid leukemias with the t(6;11) chromosome translocation.";
RL   Cancer Res. 53:5624-5628(1993).
RN   [2]
RP   SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RC   TISSUE=Fetal brain;
RX   MEDLINE=98344142; PubMed=9679199;
RA   Saito S., Matsushima M., Shirahama S., Minaguchi T., Kanamori Y.,
RA   Minami M., Nakamura Y.;
RT   "Complete genomic structure, DNA polymorphisms, and alternative
RT   splicing of the human AF-6 gene.";
RL   DNA Res. 5:115-120(1998).
RN   [3]
RP   SEQUENCE OF 337-1816 FROM N.A. (ISOFORM 2).
RA   Williams S.;
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: MAY ACT AS AN INTRACELLULAR SIGNALING COMPONENT
CC       CONTROLLED BY RAS SIGNALING PATHWAYS.
CC   -!- SUBUNIT: BINDS DIRECTLY TO ZO-1 AND OCCLUDIN.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC   -!- ALTERNATIVE PRODUCTS: 3 ISOFORMS; 1, 2 (SHOWN HERE) AND 3; ARE
CC       PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- DISEASE: INVOLVED IN ACUTE LEUKEMIAS BY A CHROMOSOMAL
CC       TRANSLOCATION T(6;11)(Q27;Q23) THAT INVOLVES MLLT4 AND MLL/HRX.
CC       THE RESULT IS A ROGUE ACTIVATOR PROTEIN.
CC   -!- SIMILARITY: CONTAINS 1 DILUTE DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 FHA DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AB011399; BAA32484.1; -.
DR   EMBL; AB011399; BAA32483.1; -.
DR   EMBL; AB011399; BAA32485.1; -.
DR   EMBL; U02478; AAC50059.1; -.
```

```
DR   EMBL; AL049698; CAB76850.1; -.
DR   HSSP; Q12923; 3PDZ.
DR   MIM; 159559; -.
DR   InterPro; IPR002710; DIL.
DR   InterPro; IPR000253; FHA_domain.
DR   InterPro; IPR001478; PDZ.
DR   InterPro; IPR000159; RA.
DR   Pfam; PF01843; DIL; 1.
DR   Pfam; PF00498; FHA; 1.
DR   Pfam; PF00595; PDZ; 1.
DR   Pfam; PF00788; RA; 2.
DR   ProDom; PD003376; DIL; 1.
DR   SMART; SM00240; FHA; 1.
DR   SMART; SM00228; PDZ; 1.
DR   SMART; SM00314; RA; 2.
DR   PROSITE; PS50106; PDZ; 1.
KW   Chromosomal translocation; Proto-oncogene; Alternative splicing.
FT   DOMAIN       36    206       RAS-INTERACTING.
FT   DOMAIN      425    491       FHA.
FT   DOMAIN      804    910       DILUTE.
FT   DOMAIN      991   1077       PDZ.
FT   DOMAIN      162    174       GLU/LYS-RICH.
FT   DOMAIN     1349   1356       POLY-PRO.
FT   DOMAIN     1371   1376       POLY-PRO.
FT   DOMAIN     1561   1571       ASP/GLU-RICH (ACIDIC).
FT   SITE         26     26       MLL FUSION POINT (IN AN ACUTE MYELOID
FT                                LEUKEMIA PATIENT).
FT   VARSPLIC   1588   1611       LQDEERRRQQQLEEMRKREAEDRA -> VKGGVLWLCPSVV
FT                                PILASACFPWG (IN ISOFORM 1).
FT   VARSPLIC   1612   1816       MISSING (IN ISOFORM 1).
FT   VARSPLIC   1666   1743       LCRPPLPRDYEPPSPSPAPGAPPPPPQRNASYLKTQVLSPD
FT                                SLFTAKFVAYNEEEEEEDCSLAGQDKYSSTRKSHGDL ->
FT                                PNSYPGSTGAAVGAHDACRDAKEKRSKSQDADSPGSSGAPE
FT                                NLTFKERQRLFSQGQDVSNKVKASRKLTELENELNTK (IN
FT                                ISOFORM 3).
FT   VARSPLIC   1744   1816       MISSING (IN ISOFORM 3).
FT   CONFLICT    373    373       G -> V (IN REF. 1).
FT   CONFLICT    391    391       P -> PGRRNHFAYYNYHTYE (IN REF. 3).
FT   CONFLICT    744    744       D -> DSSHFD (IN REF. 3).
FT   CONFLICT   1031   1031       D -> DV (IN REF. 1).
FT   CONFLICT   1408   1408       R -> P (IN REF. 1).
SQ   SEQUENCE   1816 AA;  205604 MW;  EB1FE7F04879CE8F CRC64;
```

```
  Query Match             5.6%;  Score 120.5;  DB 1;  Length 1816;
  Best Local Similarity   19.7%;  Pred. No. 3.6;
  Matches   70;  Conservative   46;  Mismatches  159;  Indels   81;  Gaps   11;

Qy      31 TAGRSPPTPTVTLGP---DCPPPPPPPPPNNNNN------------NNSKHTGHKSACV 74
           |    | ||     |   | ||||||||  :   :              ::   |  || |
Db    1328 TPAAIPATPVAVSQPIRTDLPPPPPPPPVHYAGDFDGMSMDLPLPPPPSANQIGLPSAQV 1387

Qy      75 PNMTERRRDELSEEINNLREKVMKQSEENNNLQSQVQKLTEENTTLREQVEPTPEDEDDD 134
          |:|:|       :||   : |     : | :|| :  |   : : | |
Db    1388 AAAERRKREEHQRWYE--KEKARLEEERERKRREQERKLGQMRT---QSLNPAPFSPLTA 1442

Qy     135 IELRGAAAAAAPPPP---IEEECPEDLPEKFDGNPDMLAPFMAQCQIFMEKSTRDFSVDR 191
           :::    :    |    | |  |:  |   :                   :  :  :| :
Db    1443 QQMKPEKPSTLQRPQETVIRELQPQQQPRTIE----------------RRDLQYITVSK 1485

Qy     192 VRVCFVTSMMTGRAARWASAKLERSHYLMHNYPAFMMEMKHVFEDPQRREVAKRKIRRLR 251
             :    |:      | |  |||:   ||       |:: :   | |      ::|:|
Db    1486 EELSSGDSLSPDPWKRDAKEKLEKQQQ-MHIVDMLSKEIQELQSKPDRSAEESDRLRKLM 1544

Qy     252 QGMGSVIDYSNAFQMIAQDLDWNEPALIDQYHEGLSDHIQEELSHLEVAKSLSALIGQCI 311
                           |:|     : :  :   |  :||   ::     : || | :
Db    1545 ------------------LEWQFQKRLQESKQKDEDDEEEEDDDVD-----TMLIMQRL 1580

Qy     312 HIERRLARAAAAARKPRSPPRALVLPHIASHHQVDPTEPVGGARMRLTQEEKERRRK 367
             |||       |:     |   |  :      |          |  |||: ||::
Db    1581 EAERRARLQDEERR-----RQQQLEEMRKREAED----------RARQEEERRRQE 1621
```

```
RESULT  13
CHDM_DROME
ID   CHDM_DROME     STANDARD;      PRT;  1982 AA.
AC   O97159; Q9VW50;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Chromodomain helicase-DNA-binding protein Mi-2 homolog (dMi-2).
GN   MI-2 OR CG8103.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A., FUNCTION, AND MUTAGENESIS OF GLY-737.
RX   MEDLINE=99055400; PubMed=9836641;
RA   Kehle J., Beuchle D., Treuheit S., Christen B., Kennison J.A.,
RA   Bienz M., Muller J.;
RT   "dMi-2, a hunchback-interacting protein that functions in Polycomb
RT   repression.";
RL   Science 282:1897-1900(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
CC   -!- FUNCTION: VITAL ROLE IN DEVELOPMENT. PROTEIN BINDS TO A PORTION OF
CC       HUNCHBACK (HB) PROTEIN THAT IS CRITICAL FOR REPRESSION OF BITHORAX
CC       COMPLEX (BXC) GENES. MAY ALSO FUNCTION IN POLYCOMB GROUP (PCG)
CC       REPRESSION OF HOX GENES.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- SIMILARITY: BELONGS TO THE SNF2/RAD54 HELICASE FAMILY.
CC   -!- SIMILARITY: CONTAINS 2 'CHROMO' DOMAINS.
CC   -!- SIMILARITY: CONTAINS 2 PHD-TYPE ZINC FINGERS.
CC   ---------------------------------------------------------------------
```

```
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; AF119716; AAD17276.1; -.
DR   EMBL; AE003515; AAF49099.1; ALT_SEQ.
DR   FlyBase; FBgn0013591; Mi-2.
DR   InterPro; IPR000953; Chromo.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR002464; DEAH_ATP_helicase.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR001965; PHD.
DR   InterPro; IPR000330; SNF2_N.
DR   InterPro; IPR001841; Znf_ring.
DR   Pfam; PF00385; chromo; 1.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF00628; PHD; 2.
DR   Pfam; PF00176; SNF2_N; 1.
DR   SMART; SM00298; CHROMO; 2.
DR   SMART; SM00487; DEXDc; 1.
DR   SMART; SM00490; HELICc; 1.
DR   SMART; SM00249; PHD; 2.
DR   SMART; SM00184; RING; 2.
DR   PROSITE; PS00598; CHROMO_1; FALSE_NEG.
DR   PROSITE; PS00690; DEAH_ATP_HELICASE; 1.
DR   PROSITE; PS50013; CHROMO_2; 2.
KW   DNA-binding; ATP-binding; Helicase; Nuclear protein; Repeat;
KW   Transcription regulation; Repressor; Zinc-finger.
FT   ZN_FING      380    421       PHD-TYPE.
FT   ZN_FING      440    481       PHD-TYPE.
FT   DOMAIN       488    566       CHROMO 1.
FT   DOMAIN       612    673       CHROMO 2.
FT   NP_BIND      755    762       ATP (POTENTIAL).
FT   SITE         875    878       DEAH BOX.
FT   DOMAIN        13     16       POLY-GLU.
FT   DOMAIN        70     76       POLY-LYS.
FT   DOMAIN       239    248       POLY-GLU.
FT   DOMAIN      1279   1287       POLY-GLU.
FT   DOMAIN      1672   1677       POLY-ASP.
FT   MUTAGEN      737    737       G->D: IN ALLELE MI-2-5; LARVAL LETHAL.
FT   CONFLICT     101    101       G -> A (IN REF. 1).
SQ   SEQUENCE   1982 AA;  224199 MW;  ED8E256D1AD0AC2F CRC64;


  Query Match             5.6%;  Score 120.5;  DB 1;  Length 1982;
  Best Local Similarity   18.8%;  Pred. No. 4;
  Matches   80;  Conservative   63;  Mismatches  152;  Indels  131;  Gaps   16;

Qy      27 RSEATAGRSPP---------TPTVTLGPDCPPPPPPPPPNNNNNNNSKHTGHKSACVPNM 77
           : :  |  :||         | : :: |     | | | :    :  |
Db    1529 KQDVAALEAPPTGGNVDKSATTSNSVTPATSAAPSPAPASEKGEDKDKD---------- 1577

Qy      78 TERRRDELSEEINNLREKVMKQSEENN----------------------NLQSQVQKLT 114
           :|: :|: | | : ::::  :::||:                          ::::| |
Db    1578 SEKEKDKTSAEKSEVKQE--QEAEEDKKPGDVKQENPVEEAAGDTKPSDAEVKTEVAKTE 1635

Qy     115 EENTTLREQVEPTPEDEDDDIELRGAAAAAAPPPPIEEECPEDLPEKFDGNPDMLAPFMA 174
           :  |   :|:  |: |: : |      |   |::: : :  | ||
Db    1636 PKEETKDPEVKEEPKTEEKEKEKVDDKKPIPPTTVIDDDDDDVMIVKEDGE--------- 1686

Qy     175 QCQIFMEKSTRDFSVDRVRVCFVTSMMTGRAARWASAKLE--RSHYLMHNYPAFMMEMKH 232
             :|| :     |:  |   || ||   : |   ||  :  :: :       |:
Db    1687 -----LEKPSASSPKDQKAVAAATSAATGATGKGAEDSLEVLKRKFMFNIADGGFTELHT 1741

Qy     233 VFEDPQRREVAKRKI----RR---------LRQGMGSVIDYSN--AFQMIAQDLDWNEPA 277
           :: : ::  |  |:     ||         :  | |   |  |   | :|      |||
Db    1742 LWLNEEKAAVPGREYEIWHRRHDYWLLAGIVTHGYGRWQDIQNDIRFAII------NEPF 1795

Qy     278 LIDQYHEGLSDHIQEELSHLEVAKSLSA----LIGQCIHIERRLARAAAARKPRSPPRAL 333
```

```
           :|   :|         : ||:     |    |: | : || :| |||     :
Db    1796 KMD-VGKG---------NFLEIKNKFLARRFKLLEQALVIEEQLRRAAYLNLAQ------ 1839

Qy     334 VLPHIASHHQVDPTEPVGGARMRLTQEEKERRRKLNLCLYCGTGGHYADNCPAKASKSSP 393
                   ||: |      |  : |           |   |        || |
Db    1840 ----------DPSHPAMSLNARFAEVE-----------CLAESH------QHLSKESL 1870

Qy     394 AGNSPA 399
           ||| ||
Db    1871 AGNKPA 1876


RESULT  14
HGL2_ARATH
ID   HGL2_ARATH     STANDARD;      PRT;   745 AA.
AC   P46607; Q39018;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Homeobox protein GLABRA2 (Homeobox-leucine zipper protein ATHB-10)
DE   (HD-ZIP protein ATHB-10).
GN   GL2 OR AT1G79840 OR F19K16.20.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. WASSILEWSKIJA; TISSUE=Seedling;
RX   MEDLINE=95011550; PubMed=7926739;
RA   Rerie W.G., Feldmann K.A., Marks M.D.;
RT   "The GLABRA2 gene encodes a homeo domain protein required for normal
RT   trichome development in Arabidopsis.";
RL   Genes Dev. 8:1388-1399(1994).
RN   [2]
RP   REVISIONS.
RA   Marks M.D.;
RL   Submitted (NOV-1996) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RX   MEDLINE=96407838; PubMed=8811855;
RA   di Cristina M., Sessa G., Dolan L., Linstead P., Baima S., Ruberti I.,
RA   Morelli G.;
RT   "The Arabidopsis Athb-10 (GLABRA2) is an HD-Zip protein required for
RT   regulation of root hair development.";
RL   Plant J. 10:393-402(1996).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RX   MEDLINE=21016719; PubMed=11130712;
RA   Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
RA   White O., Alonso J., Altafi H., Araujo R., Bowman C.L., Brooks S.Y.,
RA   Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
RA   Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,
RA   Dunn P., Etgu P., Feldblyum T.V., Feng J.-D., Fong B., Fujii C.Y.,
RA   Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,
RA   Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
RA   Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,
RA   Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
RA   Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziali A.,
RA   Militscher J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
RA   Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
RA   Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
RA   Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
RA   Utterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
RA   Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
RT   "Sequence and analysis of chromosome 1 of the plant Arabidopsis
RT   thaliana.";
RL   Nature 408:816-820(2000).
CC   -!- FUNCTION: REQUIRED FOR CORRECT MORPHOLOGICAL DEVELOPMENT AND
CC       MATURATION OF TRICHOMES AS WELL AS FOR NORMAL DEVELOPMENT OF SEED
```

```
CC       COAT MUCILAGE.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC   -!- TISSUE SPECIFICITY: EXPRESSED IN DEVELOPING TRICHOMES.
CC   -!- SIMILARITY: BELONGS TO THE HD-ZIP FAMILY OF HOMEOBOX PROTEINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L32873; AAC80260.1; -.
DR   EMBL; Z54356; CAA91183.1; ALT_INIT.
DR   EMBL; AC011717; AAG52245.1; -.
DR   HSSP; P22808; 1NK3.
DR   TRANSFAC; T02961; -.
DR   InterPro; IPR001356; Homeobox.
DR   InterPro; IPR002913; START.
DR   Pfam; PF00046; homeobox; 1.
DR   Pfam; PF01852; START; 1.
DR   SMART; SM00389; HOX; 1.
DR   SMART; SM00234; START; 1.
DR   PROSITE; PS00027; HOMEOBOX_1; 1.
DR   PROSITE; PS50071; HOMEOBOX_2; 1.
KW   Homeobox; DNA-binding; Nuclear protein.
FT   DOMAIN       39     60       ASP/GLU-RICH (ACIDIC).
FT   DOMAIN       70     87       ASP/GLU-RICH (ACIDIC).
FT   DNA_BIND     99    158       HOMEOBOX.
SQ   SEQUENCE   745 AA;  82988 MW;  E4AD2DAAC155D5F3 CRC64;


  Query Match             5.6%;  Score 120;  DB 1;  Length 745;
  Best Local Similarity   18.1%;  Pred. No. 1.4;
  Matches   67;  Conservative   64;  Mismatches  133;  Indels  106;  Gaps   17;

Qy      57 NNNNNNNSKHTGHKSAC------------VPNMTERRRDELSEEI-----------NNLR 93
           |   |   :   |:                |:  |::| :||:::            | |
Db      92 NKGTNKRKRKKYHRHTTDQIRHMEALFKETPHPDEKQRQQLSKQLGLAPRQVKFWFQNRR 151

Qy      94 E--KVMKQSEENNNLQSQVQKLTEENTTLREQVE------PTPEDEDDDIELRGA----- 140
             | :::  ||: |:::::|| |||  :||         |      ||: |  :
Db     152 TQIKAIQERHENSLLKAELEKLREENKAMRESFSKANSSCPNCGGGPDDLHLENSKLKAE 211

Qy     141 -----AAAAAPPPPIEEECPEDLPEKFDGNPDMLAPFMAQCQIFMEKSTRDFSVDRVRVC 195
                 ||    | |::  | :|   :  |: |                |  |:::: |:
Db     212 LDKLRAALGRTPYPLQASCSDDQEHRL-GSLDFY-------------TGVFALEKSRIA 256

Qy     196 FVTSMMTGRAARWASA-------KLERSHYLMHNYPAFMMEMKHVFEDPQRREVAKRKIR 248
           :::  |    : |::        :|    :: ||  :: |    |   |      ||
Db     257 EISNRATLELQKMATSGEPMWLRSVETGREIL-NYDEYLKE----FPQAQASSFPGRKTI 311

Qy     249 RLRQGMGSVIDYSNAFQMIAQDLD---WNE--------PALIDQYHEG-----LSDHIQ- 291
             :  | |  : :| ::    :|   | |         | :|   :|     :   ||
Db     312 EASRDAGIV--FMDAHKLAQSFMDVGQWKETFACLISKAATVDVIRQGEGPSRIDGAIQL 369

Qy     292 -------------EELSHLEVAKSLS----ALIGQCIHIE-RRLARAAAARKPRSPPRA 332
                        |:  :   : ||    |::   : :|     : |:  | |  |
Db     370 MFGEMQLLTPVVPTREVYFVRSCRQLSPEKWAIVDVSVSVEDSNTEKEASLLKCRKLPSG 429

Qy     333 LVLPHIASHH 342
           ::   :: |
Db     430 CIIEDTSNGH 439


RESULT  15
TAGB_DICDI
ID   TAGB_DICDI     STANDARD;      PRT;  1905 AA.
AC   P54683;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Prestalk-specific protein tagB precursor (EC 3.4.21.-).
GN   TAGB.
OS   Dictyostelium discoideum (Slime mold).
OC   Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX   NCBI_TaxID=44689;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AX4;
RX   MEDLINE=95262903; PubMed=7744252;
RA   Shaulsky G., Kuspa A., Loomis W.F.;
RT   "A multidrug resistance transporter/serine protease gene is required
RT   for prestalk specialization in Dictyostelium.";
RL   Genes Dev. 9:1111-1122(1995).
CC   -!- FUNCTION: INTERCELLULAR COMMUNICATION VIA TAGB MAY MEDIATE
CC       INTEGRATION OF CELLULAR DIFFERENTIATION WITH MORPHOGENESIS.
CC   -!- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO PEPTIDASE FAMILY
CC       S8; ALSO KNOWN AS THE SUBTILASE FAMILY.
CC   -!- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE ATP-BINDING
CC       TRANSPORT PROTEIN FAMILY (ABC TRANSPORTERS). MDR SUBFAMILY.
CC   -!- SIMILARITY: STRONG, TO TAGC.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U20432; AAA62212.1; -.
DR   HSSP; P13569; 1NBD.
DR   DictyDb; DD02059; tagB.
DR   InterPro; IPR003593; AAA.
DR   InterPro; IPR001140; ABC_transporter_tmem.
DR   InterPro; IPR003439; ABC_transportr.
DR   InterPro; IPR001687; ATP_GTP_A.
DR   InterPro; IPR000209; Peptidase_S8.
DR   Pfam; PF00664; ABC_membrane; 1.
DR   Pfam; PF00005; ABC_tran; 1.
DR   PRINTS; PR00723; SUBTILISIN.
DR   SMART; SM00382; AAA; 1.
DR   PROSITE; PS00136; SUBTILASE_ASP; FALSE_NEG.
DR   PROSITE; PS00137; SUBTILASE_HIS; 1.
DR   PROSITE; PS00138; SUBTILASE_SER; 1.
DR   PROSITE; PS00211; ABC_TRANSPORTER; 1.
KW   Hydrolase; Serine protease; ATP-binding; Transport; Transmembrane;
KW   Signal.
FT   SIGNAL       1     31       POTENTIAL.
FT   CHAIN       32   1905       PRESTALK-SPECIFIC PROTEIN TAGB.
FT   DOMAIN     378    700       PROTEASE.
FT   DOMAIN       ?   1905       ABC TRANSPORTER.
FT   TRANSMEM  1011   1031       POTENTIAL.
FT   TRANSMEM  1076   1096       POTENTIAL.
FT   TRANSMEM  1121   1141       POTENTIAL.
FT   TRANSMEM  1210   1230       POTENTIAL.
FT   TRANSMEM  1309   1329       POTENTIAL.
FT   TRANSMEM  1332   1352       POTENTIAL.
FT   ACT_SITE   387    387       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE   432    432       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE   695    695       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   NP_BIND   1553   1560       ATP (POTENTIAL).
FT   DOMAIN      63     67       POLY-GLN.
FT   DOMAIN      95    104       POLY-ASN.
FT   DOMAIN     107    134       POLY-ASN.
FT   DOMAIN     311    321       POLY-SER.
FT   DOMAIN     833    837       POLY-SER.
FT   DOMAIN     838    844       POLY-GLY.
FT   DOMAIN     871    876       POLY-LEU.
FT   DOMAIN    1012   1015       POLY-ILE.
FT   DOMAIN    1386   1389       POLY-GLU.
FT   DOMAIN    1398   1404       POLY-GLY.
FT   DOMAIN    1445   1450       POLY-ASN.
FT   DOMAIN    1765   1779       POLY-ASN.
```

```
FT   DOMAIN     1782   1785       POLY-SER.
FT   DOMAIN     1807   1812       POLY-PRO.
FT   DOMAIN     1813   1860       POLY-GLN.
FT   DOMAIN     1872   1878       POLY-PRO.
FT   CARBOHYD    594    594       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    621    621       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    672    672       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    747    747       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    823    823       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1172   1172       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1522   1522       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1658   1658       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1905 AA;  212518 MW;  B8E223FA8B9AE13C CRC64;


  Query Match          5.6%;  Score 120;  DB 1;  Length 1905;
  Best Local Similarity   19.9%;  Pred. No. 4.1;
  Matches   30;  Conservative   25;  Mismatches   52;  Indels   44;  Gaps    4;

Qy      26 HRSEATAGRSPPTPTVTLGPDCPPP----------PPPPPPNNNNNNNSKHTGHKSACVP 75
           : :| ::  :||    |     |||          ||||||
Db    1777 NNNEPSSSSTPPNDQPT-----PPPQEQQEQKNDQPPPPPPQEQ---------------- 1815

Qy      76 NMTERRRDELSEEINNLREKVMKQSEENNNLQSQVQKLTEENTTLREQVEPTPEDEDDDI 135
              : ::::  ::    :::  :| ::    | | |:  ::    ::| :  | | |
Db    1816 ---QEQQEQQQQQQQEQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQNDQPPNDYDQ-- 1870

Qy     136 ELRGAAAAAAPPPPIEEECPEDLPEKFDGNP 166
                  ||||:  | |       || |
Db    1871 --------VPPPPPLPSESPSPPTGNNDGQP 1893


Search completed: October  4, 2002, 15:40:09
Job time: 111 sec
```

```
OM protein - protein search, using sw model

Run on:          October  4, 2002, 15:38:18 ; Search time 31.89 Seconds
                                              (without alignments)
                                              2175.320 Million cell updates/sec

Title:           US-09-631-863A-2
Perfect score:   2139
Sequence:        1 MRNKRVLKTKKRRSGRGGQD..........DNCPAKASKSSPAGNSPAPL 401

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched:        562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters:      562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       SPTREMBL_19:*
               1:  sp_archea:*
               2:  sp_bacteria:*
               3:  sp_fungi:*
               4:  sp_human:*
               5:  sp_invertebrate:*
               6:  sp_mammal:*
               7:  sp_mhc:*
               8:  sp_organelle:*
               9:  sp_phage:*
               10:  sp_plant:*
               11:  sp_rodent:*
               12:  sp_virus:*
               13:  sp_vertebrate:*
               14:  sp_unclassified:*
               15:  sp_rvirus:*
               16:  sp_bacteriap:*
               17:  sp_archeap:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1802	84.2	342	4	Q9UPV1	Q9upv1 homo sapien
2	1703	79.6	325	4	Q96A68	Q96a68 homo sapien
3	943	44.1	231	11	Q9EQ11	Q9eq11 mus musculu
4	373	17.4	704	13	Q98SV9	Q98sv9 fugu rubrip
5	364.5	17.0	371	13	O93283	O93283 fugu rubrip
6	279	13.0	1252	3	Q96V43	Q96v43 aspergillus
7	240.5	11.2	853	3	Q00833	Q00833 fusarium ox
8	229.5	10.7	837	3	Q9HFY8	Q9hfy8 colletotric
9	223.5	10.4	400	6	Q95KK1	Q95kk1 macaca fasc
10	192.5	9.0	1418	4	Q9P2M8	Q9p2m8 homo sapien
11	189	8.8	1810	10	Q94I15	Q94i15 oryza sativ
12	185	8.6	1500	10	Q94DA0	Q94da0 oryza sativ
13	185	8.6	1862	10	Q94I04	Q94i04 oryza sativ
14	184.5	8.6	1781	10	Q9AYC0	Q9ayc0 oryza sativ
15	182.5	8.5	1473	10	Q94H22	Q94h22 oryza sativ
16	180.5	8.4	959	10	Q94I08	Q94i08 oryza sativ
17	179.5	8.4	1714	10	Q94HT9	Q94ht9 oryza sativ
18	179.5	8.4	1777	10	Q94HP9	Q94hp9 oryza sativ
19	179	8.4	2017	10	Q9AYB6	Q9ayb6 oryza sativ
20	178.5	8.3	1491	10	Q9XE26	Q9xe26 oryza sativ
21	172.5	8.1	1773	10	Q94HV6	Q94hv6 oryza sativ
22	170.5	8.0	1506	10	Q94DR3	Q94dr3 oryza sativ
23	167	7.8	1524	10	Q9LWJ0	Q9lwj0 oryza sativ
24	166.5	7.8	1475	10	Q94H45	Q94h45 oryza sativ
25	164.5	7.7	1571	10	Q94HN3	Q94hn3 oryza sativ
26	163	7.6	1521	10	Q9XEM5	Q9xem5 oryza sativ
27	163	7.6	2157	10	Q9AYB5	Q9ayb5 oryza sativ
28	162.5	7.6	406	3	Q9UVD9	Q9uvd9 alternaria
29	161.5	7.6	2162	10	Q9AYC2	Q9ayc2 oryza sativ
30	161	7.5	1220	10	Q9XHY2	Q9xhy2 oryza sativ
31	161	7.5	1521	10	Q94DL1	Q94dl1 oryza sativ
32	159.5	7.5	639	3	Q9UVC2	Q9uvc2 cladosporiu
33	159.5	7.5	963	10	Q9LWX0	Q9lwx0 oryza sativ
34	157	7.3	1524	10	Q93VN0	Q93vn0 oryza sativ
35	156	7.3	1461	10	Q94H43	Q94h43 oryza sativ
36	151	7.1	372	3	Q9Y8H2	Q9y8h2 tricholoma
37	148	6.9	1494	5	Q94885	Q94885 drosophila
38	146	6.8	1524	10	Q943A1	Q943a1 oryza sativ
39	142.5	6.7	303	4	Q9H7A0	Q9h7a0 homo sapien
40	139	6.5	1587	10	Q9SLQ0	Q9slq0 oryza sativ
41	138.5	6.5	2075	13	Q90WA4	Q90wa4 fugu rubrip
42	138	6.5	675	11	Q91YM9	Q91ym9 mus musculu
43	138	6.5	1123	11	Q9DBD5	Q9dbd5 mus musculu
44	137	6.4	498	10	Q94BM0	Q94bm0 hordeum vul
45	135	6.3	100	11	Q9DCZ3	Q9dcz3 mus musculu

ALIGNMENTS

```
RESULT   1
Q9UPV1
ID   Q9UPV1      PRELIMINARY;      PRT;   342 AA.
AC   Q9UPV1;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   KIAA1051 PROTEIN (FRAGMENT).
GN   KIAA1051.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RX   MEDLINE=99397452; PubMed=10470851;
RA   Kikuno R., Nagase T., Ishikawa K., Hirosawa M., Miyajima N.,
RA   Tanaka A., Kotani H., Nomura N., Ohara O.;
RT   "Prediction of the coding sequences of unidentified human genes. XIV.
RT   The complete sequences of 100 new cDNA clones from brain which code
RT   for large proteins in vitro.";
RL   DNA Res. 6:197-205(1999).
DR   EMBL; AB028974; BAA83003.1; -.
DR   InterPro; IPR001878; Znf_CCHC.
DR   Pfam; PF00098; zf-CCHC; 1.
DR   SMART; SM00343; ZnF_C2HC; 1.
KW   Zinc-finger.
FT   NON_TER       1      1
SQ   SEQUENCE   342 AA;  38769 MW;  DC14C265B92541EB CRC64;

  Query Match             84.2%;  Score 1802;  DB 4;  Length 342;
  Best Local Similarity   100.0%;  Pred. No. 7.8e-131;
  Matches  342;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      60 NNNNSKHTGHKSACVPNMTERRRDELSEEINNLREKVMKQSEENNNLQSQVQKLTEENTT 119
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 NNNNSKHTGHKSACVPNMTERRRDELSEEINNLREKVMKQSEENNNLQSQVQKLTEENTT 60
```

```
Qy    120 LREQVEPTPEDEDDDIELRGAAAAAAPPPPIEEECPEDLPEKFDGNPDMLAPFMAQCQIF 179
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 LREQVEPTPEDEDDDIELRGAAAAAAPPPPIEEECPEDLPEKFDGNPDMLAPFMAQCQIF 120

Qy    180 MEKSTRDFSVDRVRVCFVTSMMTGRAARWASAKLERSHYLMHNYPAFMMEMKHVFEDPQR 239
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 MEKSTRDFSVDRVRVCFVTSMMTGRAARWASAKLERSHYLMHNYPAFMMEMKHVFEDPQR 180

Qy    240 REVAKRKIRRLRQGMGSVIDYSNAFQMIAQDLDWNEPALIDQYHEGLSDHIQEELSHLEV 299
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 REVAKRKIRRLRQGMGSVIDYSNAFQMIAQDLDWNEPALIDQYHEGLSDHIQEELSHLEV 240

Qy    300 AKSLSALIGQCIHIERRLARAAAARKPRSPPRALVLPHIASHHQVDPTEPVGGARMRLTQ 359
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 AKSLSALIGQCIHIERRLARAAAARKPRSPPRALVLPHIASHHQVDPTEPVGGARMRLTQ 300

Qy    360 EEKERRRKLNLCLYCGTGGHYADNCPAKASKSSPAGNSPAPL 401
          |||||||||||||||||||||||||||||||||||||||||||
Db    301 EEKERRRKLNLCLYCGTGGHYADNCPAKASKSSPAGNSPAPL 342
```

RESULT 2
Q96A68
ID Q96A68 PRELIMINARY; PRT; 325 AA.
AC Q96A68;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE MEF3 LIKE 1 (PATERNALLY EXPRESSED GENE 10 ORF1).
GN MEF3L1 OR PEG10.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=TESTIS;
RA Nakamura Y., Furukawa Y.;
RL Submitted (SEP-2000) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=TESTIS;
RA Satoh S., Furukawa Y.;
RT "Isolation of MEF3 like gene 1.";
RL Submitted (SEP-2001) to the EMBL/GenBank/DDBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=21218929; PubMed=11318613;
RA Ono R., Kobayashi S., Wagatsuma H., Aisaka K., Kohda T.,
RA Kaneko-Ishino T., Ishino F.;
RT "A Retrotransposon-Derived Gene, PEG10, Is a Novel Imprinted Gene
RT Located on Human Chromosome 7q21.";
RL Genomics 73:232-237(2001).
DR EMBL; AB049150; BAB68387.1; -.
DR EMBL; AB049834; BAB43951.1; -.
SQ SEQUENCE 325 AA; 36965 MW; 118E4CFAF97F2A76 CRC64;

Query Match 79.6%; Score 1703; DB 4; Length 325;
Best Local Similarity 100.0%; Pred. No. 3e-123;
Matches 325; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy     77 MTERRRDELSEEINNLREKVMKQSEENNNLQSQVQKLTEENTTLREQVEPTPEDEDDDIE 136
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MTERRRDELSEEINNLREKVMKQSEENNNLQSQVQKLTEENTTLREQVEPTPEDEDDDIE 60

Qy    137 LRGAAAAAAPPPPIEEECPEDLPEKFDGNPDMLAPFMAQCQIFMEKSTRDFSVDRVRVCF 196
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 LRGAAAAAAPPPPIEEECPEDLPEKFDGNPDMLAPFMAQCQIFMEKSTRDFSVDRVRVCF 120

Qy    197 VTSMMTGRAARWASAKLERSHYLMHNYPAFMMEMKHVFEDPQRREVAKRKIRRLRQGMGS 256
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 VTSMMTGRAARWASAKLERSHYLMHNYPAFMMEMKHVFEDPQRREVAKRKIRRLRQGMGS 180

Qy    257 VIDYSNAFQMIAQDLDWNEPALIDQYHEGLSDHIQEELSHLEVAKSLSALIGQCIHIERR 316
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 VIDYSNAFQMIAQDLDWNEPALIDQYHEGLSDHIQEELSHLEVAKSLSALIGQCIHIERR 240

Qy    317 LARAAAARKPRSPPRALVLPHIASHHQVDPTEPVGGARMRLTQEEKERRRKLNLCLYCGT 376
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 LARAAAARKPRSPPRALVLPHIASHHQVDPTEPVGGARMRLTQEEKERRRKLNLCLYCGT 300

Qy    377 GGHYADNCPAKASKSSPAGNSPAPL 401
          |||||||||||||||||||||||||
Db    301 GGHYADNCPAKASKSSPAGNSPAPL 325
```

RESULT 3
Q9EQ11
ID Q9EQ11 PRELIMINARY; PRT; 231 AA.
AC Q9EQ11;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE MYELIN EXPRESSION FACTOR-3-LIKE PROTEIN (FRAGMENT).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6;
RX MEDLINE=21105984; PubMed=11158386;
RA Volff J.-N., Koerting C., Schartl M.;
RT "Ty3/Gypsy retrotransposon fossils in mammalian genomes: did they
RT evolve into new cellular functions?";
RL Mol. Biol. Evol. 18:266-270(2001).
DR EMBL; AF302691; AAG39979.1; -.
DR InterPro; IPR001878; Znf_CCHC.
DR Pfam; PF00098; zf-CCHC; 1.
DR SMART; SM00343; ZnF_C2HC; 1.
KW Zinc-finger.
FT NON_TER 1 1
SQ SEQUENCE 231 AA; 26171 MW; DEA82A2E624F3974 CRC64;

Query Match 44.1%; Score 943; DB 11; Length 231;
Best Local Similarity 77.0%; Pred. No. 5.9e-65;
Matches 181; Conservative 20; Mismatches 30; Indels 4; Gaps 2;

```
Qy    167 DMLAPFMAQCQIFMEKSTRDFSVDRVRVCFVTSMMTGRAARWASAKLERSHYLMHNYPAF 226
          ||| ||| |||:|||||||||||||:||||||||: |||||||:|||:|  |||||| ||
Db      1 DMLGPFMYQCQLFMEKSTRDFSVDRIRVCFVTSMLIGRAARWATAKLQRCTYLMHNYTAF 60

Qy    227 MMEMKHVFEDPQRREVAKRKIRRLRQGMGSVIDYSNAFQMIAQDLDWNEPALIDQYHEGL 286
          |||:|||||||||||| |||||||||| | |:|||||||||||||| ||||:||: |||
Db     61 MMELKHVFEDPQRREAAKRKIRRLRQGPGPVVDYSNAFQMIAQDLDWTEPALMDQFQEGL 120

Qy    287 SDHIQEELSHLEVAKSLSALIGQCIHIERRLARAAAARKPRSPPRALVLPHIASHHQVDP 346
          :  |: ||| |  |:|:||| |||||||||| ||| ||  |||||:|    : | ||
Db    121 NPDIRAELSRQEAPKTLAALITACIHIERRLARDAAA-KPDPSPRALVMP---PNSQTDP 176

Qy    347 TEPVGGARMRLTQEEKERRRKLNLCLYCGTGGHYADNCPAKASKSSPAGNSPAPL 401
          ||||||||||::|||||||:|||||| |||:|| |||||||:|| |||||||
Db    177 TEPVGGARMRLSKEEKERRRKMNLCLYCGNGGHFADTCPAKASKNSPPGNSPAPL 231
```

RESULT 4
Q98SV9
ID Q98SV9 PRELIMINARY; PRT; 704 AA.
AC Q98SV9;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE GAG-PROTEASE.

```
OS   Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC   Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC   Tetraodontidae; Takifugu.
OX   NCBI_TaxID=31033;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TRANSPOSON=SUSHI-SAN RETROTRANSPOSON;
RX   MEDLINE=21321259; PubMed=11428463;
RA   Butler M., Goodwin T., Simpson M., Singh M., Poulter R.;
RT   "Vertebrate LTR retrotransposons of the Tf1/sushi group.";
RL   J. Mol. Evol. 52:260-274(2001).
DR   EMBL; AF316578; AAG60684.1; -.
DR   MEROPS; A02.UPW; -.
DR   InterPro; IPR001969; Asp_protease.
DR   InterPro; IPR001995; Asp_prot_retrov.
DR   InterPro; IPR000953; Chromo.
DR   InterPro; IPR001878; Znf_CCHC.
DR   Pfam; PF00098; zf-CCHC; 1.
DR   SMART; SM00298; CHROMO; 1.
DR   SMART; SM00343; ZnF_C2HC; 1.
DR   PROSITE; PS00141; ASP_PROTEASE; UNKNOWN_1.
DR   PROSITE; PS50175; ASP_PROT_RETROV; 1.
DR   PROSITE; PS50013; CHROMO_2; 1.
KW   Protease; Zinc-finger.
SQ   SEQUENCE   704 AA;  76619 MW;  81E2E1C9AB0153B3 CRC64;

  Query Match             17.4%;  Score 373;  DB 13;  Length 704;
  Best Local Similarity   31.6%;  Pred. No. 1.7e-20;
  Matches   95;  Conservative   51;  Mismatches  127;  Indels   28;  Gaps    6;

Qy     105 NLQSQVQKLTEENTTL-----REQVEPTPEDEDDDIELRGAAAAAAPPPPIEEECPEDLP 159
           :: : :| |    | |      |||   |  |  ::  :  |||     | |   |  :|
Db      29 DIWASLQSLNTSVTDLLSLGWAEQVPSPPAPEALEVVTQRVAAA-----PWEPHVP--IP 81

Qy     160 EKFDGNPDMLAPFMAQCQIFMEKSTRDFSVDRVRVCFVTSMMTGRAARWASAKLERSHYL 219
           |:: |   : | |: || :  :     :  || :: || ::::||||:||:| ||
Db      82 ERYSGEAGVCASFLLQCSLVFDLQPLTYPGDRAKIAFVVNLLSGRAAQWATAVLENQTPA 141

Qy     220 MHNYPAFMMEMKHVFEDPQRREVAKRKIRRLRQGMGSVIDYSNAFQMIAQDLDWNEPALI 279
             ::| |  |:| ||: | :   |  :|| |||| || ||   |:::|    ||: ||
Db     142 SSSFPEFTAELKRVFDHPIQSGEAASQIRSLRQGSSSVADY---FRILAARSGWNDTALR 198

Qy     280 DQYHEGLSDHIQEELSHLEVAKSLSALIGQCIHIERRLARAAAARKPRSPPRALVLPHIA 339
             : :||:: :::||:  | :  |  ||   | :: ||           | |:|
Db     199 GVFTQGLAEALKDELATREESGDLETLISLAIQLDNRLRERHRQSNRGWPERSLESVSGR 258

Qy     340 SHHQVD--------PTEPVGGARM-----RLTQEEKERRRKLNLCLYCGTGGHYADNCPA 386
           |   |:        | | | |  |     ||:  |:  ||  |||| ||  ||:  :||
Db     259 SGAPVESSTLAPAPPQEDVDGEPMQLGPSRLSHAERRYRRSANLCLNCGQAGHFVADCPT 318

Qy     387 K 387
           :
Db     319 Q 319


RESULT   5
O93283
ID   O93283      PRELIMINARY;      PRT;   371 AA.
AC   O93283;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   GAG POLYPROTEIN.
GN   GAG.
OS   Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC   Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC   Tetraodontidae; Takifugu.
OX   NCBI_TaxID=31033;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TRANSPOSON=SUSHI-ICHI RETROTRANSPOSON;
RX   MEDLINE=98382517; PubMed=9714821;
RA   Poulter R., Butler M.;
RT   "A retrotransposon family from the pufferfish (fugu) Fugu rubripes.";
RL   Gene 215:241-249(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TRANSPOSON=SUSHI-ICHI RETROTRANSPOSON;
RA   Poulter R.T.M., Butler M.I.;
RL   Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF030881; AAC33525.1; -.
DR   InterPro; IPR001878; Znf_CCHC.
DR   Pfam; PF00098; zf-CCHC; 1.
DR   PRINTS; PR00939; C2HCZNFINGER.
DR   SMART; SM00343; ZnF_C2HC; 1.
KW   Polyprotein; Zinc-finger.
SQ   SEQUENCE   371 AA;  40540 MW;  617926FC010730AC CRC64;

  Query Match             17.0%;  Score 364.5;  DB 13;  Length 371;
  Best Local Similarity   27.6%;  Pred. No. 3.4e-20;
  Matches  101;  Conservative   62;  Mismatches  144;  Indels   59;  Gaps    8;

Qy      68 GHKSACVPNMTERRRDELSEEINNLREKVMKQ----SEENNNLQSQVQ----------K 112
           | ::  :|:  ||| :  | ::::|: :: |       | : |||  |           :
Db      12 GPRTPTLPSPLERRVEAHSAQLSSLQSELTKAFPTIQGEISELQSSSQTTSSTLSALSNQ 71

Qy     113 LTEENTTLREQVEPTPEDEDDDIELRGAAAAAAPPPPIEEECPEDL--PEKFDGNPDMLA 170
           ::   | |   ::    |         | || : |  |:      :|  |  | |: |:
Db      72 MSAMATVLASIIQKLGSDP-------GGAAPSEPSLPLSPRAEPNLASPRVFGGDFDLGK 124

Qy     171 PFMAQCQIFMEKSTRDFSVDRVRVCFVTSMMTGRAARWASAKLERSHYLMHNYPAFMMEM 230
           |: ||::         |  |  :| |:||::  :|  || | ::    |  :| ||  |
Db     125 GFLHQCELLFRHQPSRFVSDEAKVGFITSLLADKALSWAIAAVDLDPRLSSDYSAFRREF 184

Qy     231 KHVFEDPQRREVAKRKIRRLRQGMGSVIDYSNAFQMIAQDLDWNEPALIDQYHEGLSDHI 290
           | ||| |   | |  ::  |:||  || :|:  |:::| :  | | ||   |  |||: |
Db     185 KAVFEHPTYGEDAASRLLALQQGSRSVAEYTLEFRILAAESRWGETALRSAYRRGLSEAI 244

Qy     291 QEELSHLEVAKSLSALIGQCIHIERRLARAAAARKPR---------------------- 327
           ::|   :   ||: ||   : :: ||       | |
Db     245 -KDLIVRDRPSSLNELITLSLQMDERLRERRQERAQRAGGSTRQLSHRTSSAPDFSLTST 303

Qy     328 -SPPRALVLPHIASHHQVDPT-----EPVGGARMRLTQEEKERRRKLNLCLYCGTGGHYA 381
           :||     |||       |:     ||:   | ||:::|:|:| :  ||||||  ||:
Db     304 AAPP-----PHILLQSPAHPSPRVGEEPMQIGRSRLSRQEREQRLRDQLCLYCGNNGHFI 358

Qy     382 DNCPAK 387
             || :
Db     359 QACPVR 364


RESULT   6
Q96V43
ID   Q96V43      PRELIMINARY;      PRT;  1252 AA.
AC   Q96V43;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   POLYPROTEIN (FRAGMENT).
OS   Aspergillus flavus.
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC   Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX   NCBI_TaxID=5059;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NRRL 19997; TRANSPOSON=AFRTL-1;
RX   MEDLINE=95314240; PubMed=7793909;
RA   McAlpin C.E., Mannarelli B.;
RT   "Construction and characterization of a DNA probe for distinguishing
RT   strains of Aspergillus flavus.";
```

```
RL   Appl. Environ. Microbiol. 61:1068-1072(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NRRL 19997; TRANSPOSON=AFRTL-1;
RA   Okubara P.A., Tibbot B., McAlpin C.E., Hua S.-S.T.;
RT   "AfRTL-1, a retrotransposon-like element in the aflatoxin-producing
RT   fungus Aspergillus flavus.";
RL   Submitted (MAR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF362957; AAL26311.1; -.
FT   NON_TER    1252   1252
SQ   SEQUENCE   1252 AA;  144752 MW;  DEFDC5DDD8E6C4E8 CRC64;


  Query Match             13.0%;  Score 279;  DB 3;  Length 1252;
  Best Local Similarity   24.3%;  Pred. No. 5.7e-13;
  Matches   94;  Conservative   72;  Mismatches  155;  Indels   66;  Gaps   14;

Qy      58 NNNNNNSKHTGHKS-----------ACVPNMTERRRDELSEEINNLREKVMKQSEENNNL 106
           :: :::|| |  ||             |    :: :  ::: :|| :||    ::|  ||
Db       2 SSQSSSSSKKTPVKSTPPAETDSESETTVKEQLKQMKSMITQLVNNAKEK----NQEIENL 57

Qy     107 QSQVQKLTEENTTLREQVEPTPEDEDDDIELRGAAAAAAPPPPI-EEECPEDLPEKFDGN 165
           : |          | |      | :|   :|     |:||   | : : |:   | |||
Db      58 KVQ----------LGEAERIRNEQQDHIAQLDAQVGASAPKDAIGKVKLPK--AEPFDGT 105

Qy     166 PDMLAPFMAQCQIFMEKSTRDFSVDRVRVCFVTSMMTGRAARWASAKLERSHY--LMHNY 223
              |  |: |  : :  : ::   :  :| |::: : | |  |    : | :|  :  |:
Db     106 RSKLQAFLTQMNMHIHANRKNLIDEADKVIFISTHLRGAAWNWFEPYI-REYYEVVPDNW 164

Qy     224 PAFMMEM--------KHV---FEDPQRREVAKRKIRRLRQGMGSVIDYSNAFQMIAQDLD 272
                |:        ||:   | |     ||:||:::| |  ||   |:  || |   :|
Db     165 SNTTRELFTDSGNLRKHLERTFGDVDAEAVAERKLKQLYQ-RGSASTYAAEFQQIISRMD 223

Qy     273 WNEPALIDQYHEGLSDHIQEELSHLEVAKSLSALIGQCIHIERRLARA-------AAARK 325
           |||  |    |:    |:::| : ::   :|:  |   : :: |            | ||
Db     224 WNEKGLCVNLHQWSQGHVKDEFARIDRPATLNEAIDFAVKVDNRYHERLMEKRDNEAWRK 283

Qy     326 PRSPPRALVL-------------PHIASHHQVDPTEPVGGARMRLTQEEKERRRKLNLC 371
              |:                  |:     ::| ||  | ::  ::|:|:||||:  ||
Db     284 GSHRPKGQYKSNDQRERTGVKHNDPYGPKPMELDATEGQGQSK-GISQKERERRRREKLC 342

Qy     372 LYCGTGGHYADNCPAKASKSSPAGNSP 398
             ||  || : :|  | : | ||   |
Db     343 YNCGRAGHMSKDCRQKRN-SQPANRKP 368


RESULT   7
Q00833
ID   Q00833      PRELIMINARY;      PRT;   853 AA.
AC   Q00833;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   GAG POLYPROTEIN.
GN   GAG.
OS   Fusarium oxysporum.
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC   Hypocreales; mitosporic Hypocreales; Fusarium.
OX   NCBI_TaxID=5507;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=F. SP. LYCOPERSICI 42-87;
RX   MEDLINE=96132549; PubMed=8544829;
RA   Anaya N., Roncero M.I.;
RT   "Skippy, a retrotransposon from the fungal plant pathogen Fusarium
RT   oxysporum.";
RL   Mol. Gen. Genet. 249:637-647(1995).
DR   EMBL; L34658; AAA88790.1; -.
DR   InterPro; IPR001878; Znf_CCHC.
DR   Pfam; PF00098; zf-CCHC; 1.
DR   PRINTS; PR00939; C2HCZNFINGER.
DR   SMART; SM00343; ZnF_C2HC; 1.
KW   Polyprotein; Zinc-finger.
SQ   SEQUENCE   853 AA;  99422 MW;  1AEDD18CFBBA5B06 CRC64;


  Query Match             11.2%;  Score 240.5;  DB 3;  Length 853;
  Best Local Similarity   21.2%;  Pred. No. 3.3e-10;
  Matches   85;  Conservative   65;  Mismatches  157;  Indels   93;  Gaps   14;

Qy      33 GRSPPTPTVTLGPDCPPPPPPPPPNNNNNNNSKHTGHKSACVPNMTERRRDEL---SEEI 89
           |  |  |         | |   || : | :::  :   |    :  || |::|   : |:
Db      19 GHQPAAP-------ANPAPVNRPPTDQNMDDADDSSQSSD--DSEVERLREQLGNVTNEM 69

Qy      90 NNLREKV-------MKQSEENNNLQSQVQKLTEENTTLR---EQVEPTPEDEDDDIELRG 139
           | :|: :        :|:: ||| | ::  |       |   | ::|:|
Db      70 NEMRQMLEEFTALQHQQNQSNNNTQQEMYNLASAANNGRDPGEVLKPSP----------- 118

Qy     140 AAAAAAPPPPIEEECPEDLPEKFDGNPDMLAPFMAQCQIFMEKSTRDFSVDRVRVCFVTS 199
                                 || ||| |  |  |: | : |:      |  |  :| ::
Db     119 ------------------PEYFDGTPSKLPTFLTQSRAFITYYPNQFRNDSAKVMYMAG 159

Qy     200 MMTGRAARWASAKLERSHYLMHNY---PAFMME-----------------MKHVFEDPQ 238
           :   ||:|          :|::|   | : ::                    :|  |
Db     160 RLIKTAAQWFQP-------IMNDYMTNPPYKLQPRTALLFGENGRHEMEEALKMAFGTID 212

Qy     239 RREVAKRKIRRLRQGMGSVIDYSNAFQMIAQDLDWNEPALIDQYHEGLSDHIQEELSHLE 298
           :  |:|||: |:|  ||       |  :|  | |::  |:  : : | : :|:||   :
Db     213 EKGQAERKIKTLKQ-TGSASTLGVEFLQLASKLPWDQDVLMSFFFDALKEQVQQELWEKD 271

Qy     299 VAKSLSALIGQCIHIERR----LARAAAARKPRSPPRALVLPHIASHHQVDPT--EPVGG 352
           ::|   |   : |: |       | :   | |   :     :     | | :    |
Db     272 RPRTLVEYINMAVKIDDRQFAWRTRNSRGNKGRQDNKPRYHANQGRTRQTDTSYGTEAGP 331

Qy     353 ARMRLTQEEKERRRKLNLCLYCGTGGHYADNC--PAKASK 390
           : :|: :| :      |  ||  |||   |  | | ::
Db     332 MAIGMTKRDKSK----VTCYNCGKKGHYERECKNPVKTNQ 367


RESULT   8
Q9HFY8
ID   Q9HFY8      PRELIMINARY;      PRT;   837 AA.
AC   Q9HFY8;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   GAG PROTEIN.
GN   GAG.
OS   Colletotrichum gloeosporioides (Anthracnose fungus) (Glomerella
OS   cingulata).
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC   Sordariomycetes incertae sedis; Phyllachorales; Phyllachoraceae;
OC   Glomerella.
OX   NCBI_TaxID=5457;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TRANSPOSON=RETROTRANSPOSON CGRET;
RA   Zhu P., Oudemans P.V.;
RT   "A long terminal repeat retrotransposon Cgret from the phytopathogenic
RT   fungus Colletotrichum gloeosporioides on cranberry.";
RL   Curr. Genet. 0:0-0(2000).
DR   EMBL; AF264028; AAG24791.1; -.
DR   InterPro; IPR001878; Znf_CCHC.
DR   Pfam; PF00098; zf-CCHC; 1.
DR   SMART; SM00343; ZnF_C2HC; 1.
KW   Zinc-finger.
SQ   SEQUENCE   837 AA;  97738 MW;  EF1D4BC70FD55003 CRC64;


  Query Match             10.7%;  Score 229.5;  DB 3;  Length 837;
  Best Local Similarity   20.6%;  Pred. No. 2.2e-09;
  Matches   76;  Conservative   68;  Mismatches  174;  Indels   51;  Gaps    8;

Qy      68 GHKSACVP-NMTERRRDELSEEINNLR---EKVMKQSEENNNLQSQVQKLTEENTT---- 119
           |  |  || |: :   |:  :: ::     |:: || |: ||   ::: :  :|
Db      13 GSSSRQVPQNLDQELGDQKMKDDSDSEDETEELKKQLEKTNNDFKEMRDMFNQNAAALKE 72
```

```
Qy    120 LREQVEPTPEDEDDDIELRGAAAAAAPPPPIEEECPEDLPEKFDGNPDMLAPFMAQCQIF 179
          |: :     | : :   || ||   : |    |    : |  ||| |  |   : | : :
Db     73 LQGRAREAQELKKEVATLRAAANVISAPVEGRERLKLNTPATFDGTPGQLKGHLVQIRTY 132

Qy    180 MEKSTRDFSVDRVRVCFVTSMMTGRAARW-----------ASAKLERSHYLMHNYPAFM 227
              |  |  ||    : : |||  |                :    :  ::  ::
Db    133 QAFHMGTFQNDTERVVHAATFLRGRALAWFEPLQQEWLDNPVEKYSQEVRNIFFSFDGYV 192

Qy    228 MEMKHVFEDPQRREVAKRKIRRLRQGMGSVIDYSNAFQMIAQDLDWNEPALIDQYHEGLS 287
            :: :| ||  :  |:| :  |||   : : |:  |: :|  ||  : : :  :::||
Db    193 KALQSLFLDPDEKRQAERDLSNLRQNKSATL-YAAEFRRLAARLDMTDESKVFAFYQGLK 251

Qy    288 DHIQEELSHLEVAKSLSALIGQCIHI-----ERRLARAAAARKPRSPPRALVLPHIASHH 342
          | :::|:: ::   |    :  | |     |||  :    : | |  :   |    ::
Db    252 DDVKDEMAKVDTPPSFLDYVEYAIKIDNRLFERRKEKGEKRQNPNSGRKYQWQPQQRFNN 311

Qy    343 QVDPTEP---------------VGGARMRLTQ---------EEKERRRKLNLCLYCGTG 377
          | :   |               |   :  ||         :: :  :|   |  |
Db    312 QRNDNRPRNNWNSTRQSTSYGQHSGPMDLSATQKTSSWRNNNQKNDNWKKDKQCFNCNQK 371

Qy    378 GHYADNCPA 386
          || |  |||
Db    372 GHLAYECPA 380


RESULT  9
Q95KK1
ID   Q95KK1      PRELIMINARY;      PRT;   400 AA.
AC   Q95KK1;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   HYPOTHETICAL 46.6 KDA PROTEIN.
OS   Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC   Cercopithecinae; Macaca.
OX   NCBI_TaxID=9541;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=CEREBELLUM CORTEX;
RA   Osada N., Hida M., Kusuda J., Tanuma R., Iseki K., Hirai M., Terao K.,
RA   Suzuki Y., Sugano S., Hashimoto K.;
RT   "Isolation of full-length cDNA clones from macaque brain cDNA
RT   libraries.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB060816; BAB46855.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   400 AA;  46612 MW;  911D7B5ADD0EEB56 CRC64;


  Query Match             10.4%;  Score 223.5;  DB 6;  Length 400;
  Best Local Similarity   28.7%;  Pred. No. 2.6e-09;
  Matches   62;  Conservative   39;  Mismatches   84;  Indels   31;  Gaps    4;

Qy    175 QCQIFMEKSTRDFSVDRVRVCFVTSMMTGRAARWASAKLERSHYLMHNYPAFMMEMKHVF 234
          | : |:       |     || |: |  || |  || :  :     |  |:| |: |::  |
Db      2 QLETFIADHEDHFPGGAERVAFLISFFTGEAKDWAISVTQEGSPLHANFPRFLDEIRKEF 61

Qy    235 EDPQRREVAKRKIRRLRQGMGSVIDYSNAFQMIAQDLDWNEPALIDQYHEGLSDHIQEEL 294
            |    |||: ||:|:||  ::  |::||| :|| | |::  | :|: :|||:  ::||
Db     62 CGPIPPRVAKKAIRKLKQGHCTLGSYADAFQFLAQFLSWDDCHLQNQFLKGLSEFFRKEL 121

Qy    295 SHLEVAKSLSALIGQCIHIERRLARAAAARKPRSPP----RALVLPHI------------ 338
              |  || :|: |||::      | |:  |    | :: |
Db    122 LWSTEMADLDELILECVEIERKV------RVPKPIPLPGVRNILFPFAPNEEESEDEEYY 175

Qy    339 -------ASHHQVDPTEPVGGARMRLTQEEKERRRK 367
                  |  |:: | :     |||  |:| : : :
Db    176 SEDEDQEACRHRLHPKDQRN--RMRAFQQEMKEKEE 209
```

```
RESULT  10
Q9P2M8
ID   Q9P2M8      PRELIMINARY;      PRT;  1418 AA.
AC   Q9P2M8;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   KIAA1318 PROTEIN (FRAGMENT).
GN   KIAA1318.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RX   MEDLINE=20181126; PubMed=10718198;
RA   Nagase T., Kikuno R., Ishikawa K., Hirosawa M., Ohara O.;
RT   "Prediction of the coding sequences of unidentified human genes.XVI.
RT   The complete sequences of 150 new cDNA clones from brain which code
RT   for large proteins in vitro.";
RL   DNA Res. 7:65-73(2000).
DR   EMBL; AB037739; BAA92556.1; -.
FT   NON_TER       1       1
SQ   SEQUENCE   1418 AA;  147474 MW;  C3F6005F4F4374CD CRC64;


  Query Match              9.0%;  Score 192.5;  DB 4;  Length 1418;
  Best Local Similarity   20.4%;  Pred. No. 3e-06;
  Matches   78;  Conservative   57;  Mismatches  174;  Indels   73;  Gaps   10;

Qy     25 PHRSEATAGRSPPTPTVTLGPDCPPPPPPPPPNNNNNNNSKHTGHKSAC-----VPNMTE 79
          | |: |:  ||      ::    | | |                  | |     :| ||
Db   1048 PIRASASGARSTSFMRASVSGSMPMPLP--------------RATASGCGMGMSMPQMT- 1092

Qy     80 RRRDELSEEINNLREKVMKQSEENNNLQSQVQ---------KLTEENTTLREQVEPTPED 130
              : :   :   :|: |         |           | |:         :  |
Db   1093 ------ATDSRGMSTPLMRASGPGTMSTPQTAFGVMSTPEIKATDSGEASTSHINITASG 1146

Qy    131 EDDDIELRGAAAAAAPPPPIEEECPEDLPEKFDGNPDMLAPFMA---------------- 174
              :       | |||  :| |          :  || | :
Db   1147 SKPTSHMTATTPETAKPPP--KEVP---------SFGMLTPALCYLLEEQEAARGSCSVE 1195

Qy    175 -QCQIFMEKSTRDFSVDRVRVCFVTSMMTGRAARWASAKLERSHYLMHNYPAFMMEMKHV 233
           : :|  ||  : |  |  |: |: |:  | | ||   ::|    | |   :|:   : :
Db   1196 EEMEIDEEKQMKGFLDDSERMAFLVSLHLGAAERWFILQMEVGEPLSHENKSFLRRSQGI 1255

Qy    234 FEDPQRREVAKRKIRRLRQGMGSVIDYSNAFQMIAQDLDWNEPALIDQYHEGLSDHIQEE 293
          ::     ::    :   :||  ||  |:  | ::|: | |::  |  :: ||||: :  :
Db   1256 YDSLSEIDILSAVLCHPKQGQKSVRQYATDFLLLARHLSWSDAILRTRFLEGLSEAVTTK 1315

Qy    294 LSH--LEVAKSLSALIGQCIHIERRLARAAAARKPRSPPRALVLPHIASHHQVDPTEPVG 351
          :    |:|| ||  || : :: | :||       |      : |||     :     |
Db   1316 MGRIFLKVAGSLKELIDRSLYTECQLAE----EKDSPGNSSQVLPTACKRN----NEEAM 1367

Qy    352 GARMRLTQEEKERRRKLNLCLY 373
          |  :   |: :| :     | |
Db   1368 GNELSSQQQTEEHQHVSKRCYY 1389


RESULT  11
Q94I15
ID   Q94I15      PRELIMINARY;      PRT;  1810 AA.
AC   Q94I15;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   PUTATIVE RETROELEMENT.
OS   Oryza sativa (Rice).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   Ehrhartoideae; Oryzeae; Oryza.
OX   NCBI_TaxID=4530;
```

```
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NIPPONBARE;
RA   McCombie W.R.;
RT   "Rice genomic sequence.";
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NIPPONBARE;
RA   Cordum H., Johnson D., Minx P., de la Bastide M., Nascimento L.U.,
RA   Spiegel L.A., Kirchoff K.A., King L., Preston R.R., Vil M.D.,
RA   Baker J.P., Miller B., Santos L., Zutavern T., Kuit K.H.,
RA   Rodriguez S., Cunnius D.M., Balija V.S., Shah R.S., Bahret A., Bal H.,
RA   O'Shaughnessy A., Dedhia N.N., McCombie W.R.;
RL   Submitted (MAR-2001) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NIPPONBARE;
RA   Palmer L.E., Bal H., Cordum H., Johnson D., Minx P., de la Bastide M.,
RA   Nascimento L.U., Spiegel L.A., Kirchoff K.A., King L., Preston R.R.,
RA   Vil M.D., Baker J.P., Miller B., Santos L., Zutavern T., Kuit K.H.,
RA   Rodriguez S., Cunnius D.M., Balija V.S., Shah R.S., Bahret A.,
RA   O'Shaughnessy A., Dedhia N.N., McCombie W.R.;
RL   Submitted (MAY-2001) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NIPPONBARE;
RA   Palmer L.E., Bal H., Cordum H., Johnson D., Minx P., de la Bastide M.,
RA   Nascimento L.U., Spiegel L.A., Kirchoff K.A., King L., Preston R.R.,
RA   Vil M.D., Baker J.P., Miller B., Santos L., Zutavern T., Kuit K.H.,
RA   Rodriguez S., Cunnius D.M., Balija V.S., Shah R.S., Bahret A.,
RA   O'Shaughnessy A., Dedhia N.N., McCombie W.R.;
RL   Submitted (AUG-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC022352; AAK51574.1; -.
SQ   SEQUENCE   1810 AA;  205746 MW;  5AB92CD640BAB1F2 CRC64;
```

```
  Query Match             8.8%;  Score 189;  DB 10;  Length 1810;
  Best Local Similarity   20.7%;  Pred. No. 7.6e-06;
  Matches  100;  Conservative   57;  Mismatches  125;  Indels  202;  Gaps   24;

Qy       5 RVLKTKKR-RSGRGGQDPGLHPHRSEATAGRSPPTPTVTLGPDCPPPPPPPPPNNNNNNN 63
           ::: |::   :| | |  | :|:        :|| |        |||||||||: |
Db     304 QMVTTRRNVNTGEGNQPEGSNPNHQ-----GNPPPP--------PPPPPPPPPDTN---- 346

Qy      64 SKHTGHKSACVPNMTERRRDELSEEINNLREKVMKQSEENNNLQSQVQKLTEENTTLREQ 123
             | :  :       |::: | :                                |
Db     347 --------AILTQI-------LAQQANMM-----------------------TAFLHH 366

Qy     124 VEPTPEDEDDDIELRGAAAAAAPPPPIEEECPEDL---PEKF--DGNP-DMLAPFMA--- 174
           ::  |:              | ||||   :  | |   |  |    || | |    |
Db     367 LQNPPQQN-----------APPPPPQHSKLAEFLRIRPPTFSSSNNPVDALDWLHAVGK 414

Qy     175 -----QCQIFMEKSTRDFSVDRVRVCFVTSMMTGRAARW-----------------ASA 211
                ||             |  :| |    : | |: |                   :|
Db     415 KLDTVQCS-----------DEEKVVFAAHQLQGPASLWWDHFQATQPEGQPITWTRFTA 462

Qy     212 KLERSHYLMHNYPAFMMEMKHVFEDPQRREVAKRKIRRLRQGMGSVIDYSNAFQMIA--- 268
             |:|      || :: :|             ||: | |:||  ||::| : |  :|
Db     463 AFRRTH-----VPAGVVALK-----------KREFRELKQGNRSVMEYLHEFNNLARYA 505

Qy     269 -QDLDWNEPALIDQYHEGLSDHIQEELSHLEVA---KSLSALIGQCIHIE---------- 314
            :|:  :|     :  |     :  |||   |:        |: : | :|
Db     506 PEDVREDE-----EKQEKFLAGMDPELSVRLVSGDYPDFQRLVDKSIRLEAKHKELESHK 560

Qy     315 RRLA------------------------RAAAARKPRSPPRALVLPHIASHHQVD---- 345
           ||||                        :    ::||| ||   :  :    |
Db     561 RRLANFRNQQGANQRVRYTNPYPGGSSSQQQQQQQQPRSAPRPQFVVRVPQPQQQQNQQG 620

Qy     346 --------PT-EPVGGARMRLTQEEKERRRKLNLCLYCGTGGHYADNCP---AKASKSSP 393
                  || :|| | |         : : :  ||  |   ||:|| ||    :  :: |
Db     621 TRAPRPPTPTVQPVQGRR--------DAQGQQRLCFNCFEPGHFADKCPKPRRQQGQAPP 672
```

```
Qy     394 AGNS 397
             |:
Db     673 RSNN 676
```

```
RESULT  12
Q94DA0
ID   Q94DA0      PRELIMINARY;      PRT;  1500 AA.
AC   Q94DA0;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   PUTATIVE POLYPROTEIN.
GN   P0439E11.5.
OS   Oryza sativa (Rice).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   Ehrhartoideae; Oryzeae; Oryza.
OX   NCBI_TaxID=4530;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. NIPPONBARE;
RA   Sasaki T., Matsumoto T., Yamamoto K.;
RT   "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
RT   clone:P0439E11.";
RL   Submitted (FEB-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AP003315; BAB61255.1; -.
KW   Polyprotein.
SQ   SEQUENCE   1500 AA;  171517 MW;  5DE27E21F9ABE861 CRC64;
```

```
  Query Match             8.6%;  Score 185;  DB 10;  Length 1500;
  Best Local Similarity   20.3%;  Pred. No. 1.2e-05;
  Matches   98;  Conservative   53;  Mismatches  125;  Indels  206;  Gaps   21;

Qy       6 VLKTKKR-RSGRGGQDPGLHPHRSEATAGRSPPTPTVTLGPDCPPPPPPPPPNNNNNNNS 64
           :: |::   :| | |  | :|:                    |   ||||||||: |
Db       1 MVSTRRNVNTGEGNQPEGSNPN-----------------PQGNPPPPPPPPDTN----- 37

Qy      65 KHTGHKSACVPNMTERRRDELSEEINNLREKVMKQSEENNNLQSQVQKLTEENTTLREQV 124
                                          | : : :|:   |     :| | ::|
Db      38 -------------------------AILTQILAQQANMMNAFLHHLQNLPQQN------- 65

Qy     125 EPTPEDEDDDIELRGAAAAAAPPPPIEEECPEDL---PEKF--DGNP-DMLAPFMA---- 174
                           | ||||   :  | |   |  |    || | |    |
Db      66 -------------------APPPPPQHSKLAEFLRIRPPTFSSSNNPVDALDWLHAVGKK 106

Qy     175 ----QCQIFMEKSTRDFSVDRVRVCFVTSMMTGRAARW-----------------ASAK 212
                ||             |  :| |    : | |: |                   :|
Db     107 LDTVQCS-----------DEEKVIFAAHQLQGPASLWWDHFQATQPEGQPITWARFTAA 154

Qy     213 LERSHYLMHNYPAFMMEMKHVFEDPQRREVAKRKIRRLRQGMGSVIDYSNAFQMIA---- 268
             |:|      || :: :|             ||: | |:||  ||::| : |  :|
Db     155 FRRTH-----VPAGVVALK-----------KREFRELKQGNRSVMEYLHEFNNLARYAP 197

Qy     269 QDLDWNEPALIDQYHEGLSDHIQEELSHLEVA---KSLSALIGQCIHIE----------R 315
           :|:  :|     :  |     :  |||   |:        |: : | :|         |
Db     198 EDVREDE-----EKQEKFLAGMDPELSVRLVSGDYPDFQRLVDKSIRLEAKHKELESHKR 252

Qy     316 RLA------------------------RAAAARKPRSPPRALVLPHIASHHQVD------ 345
           |||                        :    ::||| ||   :  :    |
Db     253 RLANFRNQQGANQRVRYTNPYPGGSSSQQQQQQQPRSAPRPQFVVRVPQQQQQQNQQGTR 312

Qy     346 ------PT-EPVGGARMRLTQEEKERRRKLNLCLYCGTGGHYADNCP---AKASKSSPAG 395
                  || :|| | |         : : :  ||  |   ||:|| ||    :  :: |
Db     313 APRPPTPTVQPVQGRR--------DAQGQQRLCFNCFEPGHFADKCPKPRRQQGQAPPRS 364

Qy     396 NS 397
           |:
Db     365 NN 366
```

```
RESULT  13
```

```
Q94I04
ID   Q94I04      PRELIMINARY;      PRT;  1862 AA.
AC   Q94I04;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   PUTATIVE RETROELEMENT.
OS   Oryza sativa (Rice).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   Ehrhartoideae; Oryzeae; Oryza.
OX   NCBI_TaxID=4530;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NIPPONBARE;
RA   McCombie W.R.;
RT   "Rice genomic sequence.";
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NIPPONBARE;
RA   Cordum H., Johnson D., Minx P., de la Bastide M., Nascimento L.U.,
RA   Spiegel L.A., Kirchoff K.A., King L., Preston R.R., Vil M.D.,
RA   Baker J.P., Miller B., Santos L., Zutavern T., Kuit K.H.,
RA   Rodriguez S., Cunnius D.M., Balija V.S., Shah R.S., Bahret A., Bal H.,
RA   O'Shaughnessy A., Dedhia N.N., McCombie W.R.;
RL   Submitted (MAR-2001) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NIPPONBARE;
RA   Palmer L.E., Bal H., Cordum H., Johnson D., Minx P., de la Bastide M.,
RA   Nascimento L.U., Spiegel L.A., Kirchoff K.A., King L., Preston R.R.,
RA   Vil M.D., Baker J.P., Miller B., Santos L., Zutavern T., Kuit K.H.,
RA   Rodriguez S., Cunnius D.M., Balija V.S., Shah R.S., Bahret A.,
RA   O'Shaughnessy A., Dedhia N.N., McCombie W.R.;
RL   Submitted (MAY-2001) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NIPPONBARE;
RA   Palmer L.E., Bal H., Cordum H., Johnson D., Minx P., de la Bastide M.,
RA   Nascimento L.U., Spiegel L.A., Kirchoff K.A., King L., Preston R.R.,
RA   Vil M.D., Baker J.P., Miller B., Santos L., Zutavern T., Kuit K.H.,
RA   Rodriguez S., Cunnius D.M., Balija V.S., Shah R.S., Bahret A.,
RA   O'Shaughnessy A., Dedhia N.N., McCombie W.R.;
RL   Submitted (AUG-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC022352; AAK51585.1; -.
SQ   SEQUENCE   1862 AA;  211396 MW;  291B106D83539AB4 CRC64;
```

```
  Query Match             8.6%;  Score 185;  DB 10;  Length 1862;
  Best Local Similarity   20.2%;  Pred. No. 1.6e-05;
  Matches   97;  Conservative   56;  Mismatches  133;  Indels  194;  Gaps   22;

Qy       5 RVLKTKKRRSGRGGQDPGLHPHRSEATAGRSPPTPTVTLGPDCPPPPPPPPPNNNNNNNS 64
           |:: |::  :   |  |    |  :     :|| |        |||||||||: |
Db     356 RMVTTRRNANTGDGNQPECSNHNHQG----NPPPP--------PPPPPPPPPDTN----- 398

Qy      65 KHTGHKSACVPNMTERRRDELSEEINNLREKVMKQSEENNNLQSQVQKLTEENTTLREQV 124
                 |  :   :       |::: | :                          |     :
Db     399 -------AILTQI-------LAQQANMM------------------------TAFLHHL 419

Qy     125 EPTPEDEDDDIELRGAAAAAAPPPPIEEECPEDL---PEKF--DGNP-DMLAPFMA---- 174
           :  |:              | ||||   :  | |   |  |    || | |    |
Db     420 QNPPQQN-----------APPPPPQHSKLAEFLRIRPPTFSSSNNPVDALDWLHAVGKK 467

Qy     175 ----QCQIFMEKSTRDFSVDRVRVCFVTSMMTGRAARW-----------------ASAK 212
               ||             |  :| |    : | |: |                   :|
Db     468 LGTVQCS-----------DEEKVIFAAHQLQGPASLWWDHFQATQPEGQPITWARFTAA 515

Qy     213 LERSHYLMHNYPAFMMEMKHVFEDPQRREVAKRKIRRLRQGMGSVIDYSNAFQMIA---- 268
             |:|      || :: :|           ||: | |:||  ||::| : |  :|
Db     516 FRRTH-----VPAGVVALK-----------KREFRELKQGNRSVMEYLHEFNNLARYAP 558
```

```
Qy     269 QDLDWNEPALIDQYHEGLSDHIQEELSHLEVA---KSLSALIGQCIHIE----------R 315
           :|:  :|     :  |     :  |||   ::        |: : | :|          |
Db     559 EDVREDE-----EKQEKFLAGMDPELSVRLISGDYPDFQRLVDKSIRLEAKHKELESHKR 613

Qy     316 RLA------------------------RAAAARKPRSPPRALVLPHIASHHQVD----- 345
           |||                        :   ::|||  ||  :  :    |
Db     614 RLANFRNQQGANQRVRYTHPYPGGSSSQQQQQQQQPRSAPRPQFVVRVPQPQQQQNQQGT 673

Qy     346 -----PTEPVGGARMRLTQEEKERRRKLNLCLYCGTGGHYADNCP---AKASKSSPAGNS 397
                ||  |   : |   : ::|     ||  |   ||:|| ||    :  :: |  |:
Db     674 RAPRPPTPTVQPGQGRRDAQGQQR-----LCFNCFEPGHFADKCPKPRRQQGQAPPRSNN 728
```

```
RESULT  14
Q9AYC0
ID   Q9AYC0      PRELIMINARY;      PRT;  1781 AA.
AC   Q9AYC0;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   POLYPROTEIN.
GN   OSJNBA0094H10.15.
OS   Oryza sativa (Rice).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   Ehrhartoideae; Oryzeae; Oryza.
OX   NCBI_TaxID=4530;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NIPPONBARE;
RA   Spiegel L.A., Nascimento L.U., de la Bastide M., Kirchoff K.A.,
RA   King L., Preston R.R., Vil M.D., Baker J.P., Miller B., Zutavern T.,
RA   Rodriguez S., Santos L., Kuit K.H., Cunnius D.M., Balija V.S.,
RA   Shah R.S., Bahret A., Bal H.P., O'Shaughnessy A., Dedhia N.N.,
RA   McCombie W.R.;
RT   "Genomic Sequence For Oryza sativa, Nipponbare Strain, Chromosome X,
RT   Clone OSJNBa0058E19, complete sequence.";
RL   Submitted (FEB-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC080019; AAK13118.1; -.
DR   InterPro; IPR001969; Asp_protease.
DR   InterPro; IPR001584; Rve.
DR   InterPro; IPR000477; RVTse.
DR   InterPro; IPR001878; Znf_CCHC.
DR   Pfam; PF00665; rve; 1.
DR   Pfam; PF00078; rvt; 1.
DR   Pfam; PF00098; zf-CCHC; 1.
DR   SMART; SM00343; ZnF_C2HC; 1.
DR   PROSITE; PS00141; ASP_PROTEASE; UNKNOWN_1.
KW   RNA-directed DNA polymerase; Zinc-finger.
SQ   SEQUENCE   1781 AA;  202845 MW;  9984A707A125FE8A CRC64;
```

```
  Query Match             8.6%;  Score 184.5;  DB 10;  Length 1781;
  Best Local Similarity   20.1%;  Pred. No. 1.6e-05;
  Matches   95;  Conservative   56;  Mismatches  141;  Indels  181;  Gaps   19;

Qy       5 RVLKTKKRRSGRGGQDPGLHPHRSEATAGRSPPTPTVTLGPDCPPPPPPPPPNNNNNNNS 64
           |:: |::  :   |  |    | :      ||| |        |||||||||: |
Db     285 RMVMTRRNAATSDGNQPEDSNHNNHG----SPPPP--------PPPPPPPPPDTN----- 327

Qy      65 KHTGHKSACVPNMTERRRDELSEEINNLREKVMKQSEENNNLQSQVQKLTEENTTLREQV 124
                 | :  :  :: : ::  :::|       |:   :|
Db     328 -------AILTQILAQQANMMTAFLHHL------QNPPQHN------------------- 355

Qy     125 EPTPEDEDDDIELRGAAAAAAPPPPIEEECPEDL---PEKF--DGNP-DMLAPFMA---- 174
                               | ||||   :  | |   |  |    || | |    |
Db     356 ------------------APPPPPQHSKLAEFLRICPPTFSSSNNPVDALDWLHAVGKK 396

Qy     175 ----QCQIFMEKSTRDFSVDRVRVCFVTSMMTGRAARW-----------------ASAK 212
               ||             |  :| |    : | |: |                   :|
Db     397 LDTVQCS-----------DEEKVIFAAHQLQGPASLWWDHFQATQPEGQPITWARFTAA 444

Qy     213 LERSHYLMHNYPAFMMEMKHVFEDPQRREVAKRKIRRLRQGMGSVIDYSNAFQMIA---- 268
```

```
          |:|      || :: :|           ||: | |:||  ||::| : |  :|
Db    445 FRRTH-----VPAGVVALK-----------KREFRELKQGNRSVMEYLHEFNNLARYAP 487

Qy    269 QDLDWNEPALIDQYHEGLSDHIQEELSHLEVA---KSLSALIGQCIHIE----------R 315
          :|:  :|     :  |     :  |||   |:        |: : | :|          |
Db    488 EDVREDE-----EKQEKFLAGMDPELSVCLVSGDYPDFQRLVDKSIRLEAKHKELESHKR 542

Qy    316 RLA------------------------RAAAARKPRSPPRALVLPHIASHHQVDPTEPV 350
          |||                        :    ::||| ||   :  :    |    :
Db    543 RLANFHNQQGANQRVRYTNPYPGGSSSQQQQQQQQPRSAPRPQFVVRVPQPQQEQNQQGT 602

Qy    351 GGAR-----MRLTQEEKERRRKLNLCLYCGTGGHYADNCPAKASKSSPAGNSP 398
            |     ::  |  :: :    ||  |   ||:|| ||    :   |   |
Db    603 RAPRPPPPTVQPGQGRRDAQGPQRLCFNCFEPGHFADKCPKPRRQQGQAPPRP 655


RESULT  15
Q94H22
ID   Q94H22      PRELIMINARY;      PRT;  1473 AA.
AC   Q94H22;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   PUTATIVE POLYPROTEIN.
OS   Oryza sativa (Rice).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   Ehrhartoideae; Oryzeae; Oryza.
OX   NCBI_TaxID=4530;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. NIPPONBARE;
RA   Buell C.R., Yuan Q., Ouyang S., Moffat K.S., Hill J.N., Gansberger K.,
RA   Brenner M., Burgess S., Hance M., Shvartsbeyn M., Tsitrin T.,
RA   Riggs F., Hsiao J., Zismann V., Blunt S., Pai G., VanAken S.E.,
RA   Utterback T.R., Feldblyum T.V., Quackenbush J., Salzberg S.L.,
RA   White O., Fraser C.M.;
RT   "Oryza sativa chromosome 3 BAC OSJNBa0077G22 genomic sequence.";
RL   Submitted (AUG-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC084831; AAK52169.1; -.
KW   Polyprotein.
SQ   SEQUENCE   1473 AA;  169336 MW;  9978B784571B22B4 CRC64;


  Query Match             8.5%;  Score 182.5;  DB 10;  Length 1473;
  Best Local Similarity   21.6%;  Pred. No. 1.9e-05;
  Matches   88;  Conservative   59;  Mismatches  125;  Indels  135;  Gaps   19;

Qy     55 PPNNNNNNNSKHTGHKSACVPNMTERRRDELSEEINNLREKVMKQSEENNNLQSQVQKLT 114
          | ||||||| ::        | : :    : :: :| : :::  |  : ||       ||
Db    265 PNNNNNNNNGEN--------PTLAQVLAQQ-AQLMNMMMQQLQNQQNQGNNHAPPQNKLV 315

Qy    115 EENTTLREQVEPTPEDEDDDIELRGAAAAAAPPPPIEEECPEDLPEKFDGNPDMLAPFMA 174
          |                    : :|    ::   |    :    : :| |         :
Db    316 E------------------FLRVRPPTFSSTTNPVEAGDWLHAIEKKLD---------LL 348

Qy    175 QCQIFMEKSTRDFSVDRVRVCFVTSMMTGRAARW-----------------ASAKLERS 216
          ||             |: :| | :  : | |: |                  :|   ::
Db    349 QC------------TDQEKVSFASHQLHGPASEWWDHFRLNRTTAEPITWLEFTAAFRKT 396

Qy    217 HYLMHNYPAFMMEMKHVFEDPQRREVAKRKIRRLRQGMGSVIDYSNAFQMIA----QDLD 272
          |      |: :: :|            |:: | | ||  || :| : |  :|    :|:
Db    397 H-----IPSGVVSLK-----------KKEFRSLTQGSRSVTEYLHEFNRLARYAPEDVR 439

Qy    273 WNEPALIDQYHEGLSDHIQEELSHLEVA---KSLSALIGQCIHIE----------RRLAR 319
           :|    ::: |||:|    |||:  :         |: : |  |          ||:|
Db    440 TDEQRQ-ERFLEGLND----ELSYPLMTGDYHDFQKLVDKAIRQEDKYNRMEQKKRRIAH 494

Qy    320 AAAARKPRSPPRALV----LPHIASHHQVDP----------------TEPVGGARMRLT 358
            | :     ||  :    :|   |   | |                  ||  :    |
Db    495 FKAQQGNSQRPRLTLGPQSMPQGGSSSVVRPQRQFFNNNAGNNIRNQAPRPVAAS----T 550

Qy    359 QEE---KERRRKLNLCLYCGTGGHYADNCP-AKASKSSPA-GNSPAP 400
```

```
          |::   ||:  |  :|  ||  ||||| ||  :  |  ||  || ||
Db    551 QQQPAKKEQGIKPGVCFNCGDPGHYADKCPKPRRVKVVPAQSNSTAP 597


Search completed: October  4, 2002, 15:40:49
Job time: 151 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: October 4, 2002, 15:38:18 ; Search time 14.69 Seconds
(without alignments)
666.757 Million cell updates/sec

Title: US-09-631-863A-2
Perfect score: 2139
Sequence: 1 MRNKRVLKTKKRRSGRGGQD.....DNCPAKASKSSPAGNSPAPL 401

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	159	7.4	146	4	US-09-209-525-2	Sequence 2, Appli
2	155	7.2	126	4	US-09-209-525-45	Sequence 45, Appl
3	149.5	7.0	110	4	US-09-209-525-52	Sequence 52, Appl
4	125	5.8	905	2	US-08-574-959A-9	Sequence 9, Appli
5	125	5.8	905	4	US-09-357-014-9	Sequence 9, Appli
6	125	5.8	1135	2	US-08-574-959A-7	Sequence 7, Appli
7	125	5.8	1135	4	US-09-357-014-7	Sequence 7, Appli
8	120	5.6	80	4	US-09-209-525-54	Sequence 54, Appl
9	118.5	5.5	2293	4	US-09-368-590-2	Sequence 2, Appli
10	118	5.5	834	3	US-08-539-205A-6	Sequence 6, Appli
11	115	5.4	1651	4	US-09-540-245A-18	Sequence 18, Appl
12	114	5.3	640	3	US-09-026-343-2	Sequence 2, Appli
13	112	5.2	1612	3	US-08-545-860D-48	Sequence 48, Appl
14	112	5.2	1612	5	PCT-US94-04496-48	Sequence 48, Appl
15	110.5	5.2	538	4	US-09-370-368-9	Sequence 9, Appli
16	109.5	5.1	1939	4	US-09-310-187A-1	Sequence 1, Appli
17	107	5.0	418	2	US-09-026-587-1	Sequence 1, Appli
18	107	5.0	418	2	US-09-227-420-1	Sequence 1, Appli
19	107	5.0	1829	4	US-09-157-420-1	Sequence 1, Appli
20	106.5	5.0	1099	4	US-09-442-100-2	Sequence 2, Appli
21	106	5.0	1872	1	US-08-188-582-14	Sequence 14, Appl
22	106	5.0	1872	1	US-08-646-715-14	Sequence 14, Appl
23	106	5.0	1886	4	US-08-938-105-3	Sequence 3, Appli
24	105.5	4.9	657	4	US-09-370-368-7	Sequence 7, Appli
25	104.5	4.9	393	2	US-09-026-587-3	Sequence 3, Appli
26	104.5	4.9	393	2	US-09-227-420-3	Sequence 3, Appli
27	104	4.9	2414	1	US-08-227-536-2	Sequence 2, Appli
28	104	4.9	2414	5	PCT-US95-04682-2	Sequence 2, Appli
29	104	4.9	3969	4	US-08-061-376-5	Sequence 5, Appli
30	102.5	4.8	413	3	US-09-089-879-7	Sequence 7, Appli
31	102.5	4.8	1137	1	US-08-369-043-2	Sequence 2, Appli
32	102.5	4.8	1719	2	US-08-459-568-4	Sequence 4, Appli
33	102.5	4.8	1719	2	US-08-399-411-4	Sequence 4, Appli
34	102.5	4.8	1719	3	US-08-516-859A-4	Sequence 4, Appli
35	102	4.8	2843	1	US-07-741-940-2	Sequence 2, Appli
36	102	4.8	2843	1	US-08-289-548A-2	Sequence 2, Appli
37	102	4.8	2843	1	US-08-452-654-2	Sequence 2, Appli
38	102	4.8	2843	2	US-08-370-235A-2	Sequence 2, Appli
39	101.5	4.7	1103	3	US-09-162-373-1	Sequence 1, Appli
40	101.5	4.7	1103	4	US-09-467-946-1	Sequence 1, Appli
41	101	4.7	1162	2	US-08-728-323A-2	Sequence 2, Appli
42	100.5	4.7	1388	4	US-09-572-191-2	Sequence 2, Appli
43	100.5	4.7	2842	1	US-07-741-940-7	Sequence 7, Appli
44	100.5	4.7	2842	1	US-08-289-548A-7	Sequence 7, Appli
45	100.5	4.7	2842	1	US-08-452-654-7	Sequence 7, Appli

ALIGNMENTS

RESULT 1
US-09-209-525-2
; Sequence 2, Application US/09209525
; Patent No. 6303770
; GENERAL INFORMATION:
; APPLICANT: Lok, Si
; APPLICANT: Conklin, Darrell C.
; APPLICANT: Parrish, Julia E.
; TITLE OF INVENTION: Mammalian Alpha Helical Protein-1
; FILE REFERENCE: 97-71
; CURRENT APPLICATION NUMBER: US/09/209,525
; CURRENT FILING DATE: 1998-12-10
; NUMBER OF SEQ ID NOS: 56
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 146
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-209-525-2

Query Match 7.4%; Score 159; DB 4; Length 146;
Best Local Similarity 27.4%; Pred. No. 3.3e-07;
Matches 45; Conservative 21; Mismatches 68; Indels 30; Gaps 2;

```
Qy      83 DELSEEINNLREKVMKQSEENNNLQSQVQKLTEENTTLREQVEPTPEDEDDDIELRGAAA 142
           |||   :: |  :    | ||: | |:: |  |  :|  || |
Db       3 DELVLLLHALLMRHRALSIENSQLMEQLRLLVCERASLLRQVRP---------------- 46

Qy     143 AAAPPPPIEEECPEDLPEKFDGNPDMLAPFMAQCQIFMEKSTRDFSVDRVRVCFVTSMMT 202
                 ||   || |:|    |  |: |   :|  :   |  | ::| |: |::|
Db      47 ---------PSCPVPFPETFNGESSRLPEFIVQTASYMLVNENRFCNDAMKVAFLISLLT 97

Qy     203 GRAARWASAKLERSHYLMHNYPAFMMEMKHVF-----EDPQRRE 241
           | |  |    :|    :: :| ||: |||  |     ||    |
Db      98 GEAEEWVVPYIEMDSPILGDYRAFLDEMKQCFGWDDDEDDDDEE 141
```

RESULT 2
US-09-209-525-45
; Sequence 45, Application US/09209525
; Patent No. 6303770
; GENERAL INFORMATION:
; APPLICANT: Lok, Si
; APPLICANT: Conklin, Darrell C.
; APPLICANT: Parrish, Julia E.
; TITLE OF INVENTION: Mammalian Alpha Helical Protein-1
; FILE REFERENCE: 97-71
; CURRENT APPLICATION NUMBER: US/09/209,525
; CURRENT FILING DATE: 1998-12-10

```
;  NUMBER OF SEQ ID NOS: 56
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 45
;   LENGTH: 126
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-209-525-45


  Query Match             7.2%;  Score 155;  DB 4;  Length 126;
  Best Local Similarity   27.6%;  Pred. No. 6.2e-07;
  Matches   40;  Conservative   18;  Mismatches   57;  Indels   30;  Gaps    2;

Qy    102 ENNNLQSQVQKLTEENTTLREQVEPTPEDEDDDIELRGAAAAAAPPPPIEEECPEDLPEK 161
          ||: |  |:: |  |  :|  || |                              ||   ||
Db      2 ENSQLMEQLRLLVCERASLLRQVRP------------------------PSCPVPFPET 36

Qy    162 FDGNPDMLAPFMAQCQIFMEKSTRDFSVDRVRVCFVTSMMTGRAARWASAKLERSHYLMH 221
          |:|    |  |: |   :|  :   |  | ::| |: |::|| |  |    :|    ::
Db     37 FNGESSRLPEFIVQTASYMLVNENRFCNDAMKVAFLISLLTGEAEEWVVPYIEMDSPILG 96

Qy    222 NYPAFMMEMKHVF-----EDPQRRE 241
          :| ||: |||  |     ||    |
Db     97 DYRAFLDEMKQCFGWDDDEDDDDEE 121


RESULT   3
US-09-209-525-52
; Sequence 52, Application US/09209525
; Patent No. 6303770
; GENERAL INFORMATION:
;  APPLICANT: Lok, Si
;  APPLICANT: Conklin, Darrell C.
;  APPLICANT: Parrish, Julia E.
;  TITLE OF INVENTION: Mammalian Alpha Helical Protein-1
;  FILE REFERENCE: 97-71
;  CURRENT APPLICATION NUMBER: US/09/209,525
;  CURRENT FILING DATE: 1998-12-10
;  NUMBER OF SEQ ID NOS: 56
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 52
;   LENGTH: 110
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-209-525-52


  Query Match             7.0%;  Score 149.5;  DB 4;  Length 110;
  Best Local Similarity   27.3%;  Pred. No. 1.6e-06;
  Matches   36;  Conservative   18;  Mismatches   53;  Indels   25;  Gaps    1;

Qy    103 NNNLQSQVQKLTEENTTLREQVEPTPEDEDDDIELRGAAAAAAPPPPIEEECPEDLPEKF 162
          |: |  |:: |  |  :|  || |                              ||   || |
Db      1 NSQLMEQLRLLVCERASLLRQVRP------------------------PSCPVPFPETF 35

Qy    163 DGNPDMLAPFMAQCQIFMEKSTRDFSVDRVRVCFVTSMMTGRAARWASAKLERSHYLMHN 222
          :|    |  |: |   :|  :   |  | ::| |: |::|| |  |    :|    :: :
Db     36 NGESSRLPEFIVQTASYMLVNENRFCNDAMKVAFLISLLTGEAEEWVVPYIEMDSPILGD 95

Qy    223 YPAFMMEMKHVF 234
          | ||: |||  |
Db     96 YRAFLDEMKQCF 107


RESULT   4
US-08-574-959A-9
; Sequence 9, Application US/08574959A
; Patent No. 5962224
;  GENERAL INFORMATION:
;    APPLICANT:  Jaekyoon Shin, Insil Joung, Ratna K. Vadlamudi
;    APPLICANT:  and Jack L. Strominger
;    TITLE OF INVENTION:  p62 POLYPEPTIDES, RELATED POLYPEPTIDES
;    TITLE OF INVENTION:  AND USES THEREFOR
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LAHIVE & COCKFIELD
;      STREET:  60 State Street, Suite 510
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02109-1875
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/574,959A
;      FILING DATE:  19-DEC-95
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Mandragouras, Amy E.
;      REGISTRATION NUMBER:  36,207
;      REFERENCE/DOCKET NUMBER:  DFN-008
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)227-7400
;      TELEFAX:  (617)227-5941
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  905 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-574-959A-9


  Query Match             5.8%;  Score 125;  DB 2;  Length 905;
  Best Local Similarity   20.5%;  Pred. No. 0.0055;
  Matches   45;  Conservative   22;  Mismatches   75;  Indels   78;  Gaps    7;

Qy     22 GLHPHRSEATAGRSPPTPTVTLGP--------------------DCPPPPPPPPP---- 56
          || |       :| :|| |    ||                        |||||||||
Db    560 GLPPLPPPPPSGATPP-PIAPTGPPTASPPVPAKEEPEELPAAPGPLPPPPPPPPPVPGP 618

Qy     57 ----------------------------NNNNNNNSKHTGHKSACVPNMTERRRDE-- 84
                                        |:::   :  | :       |   :|
Db    619 VXLPPPQLVPEGTPGGGGPPALEEDLTVININSSDEEEEEEGEEEEEEEEEEEEEEEE 678

Qy     85 ------LSEEINNLREKVMKQSEENNNLQSQVQKLTEENTTLREQVEPTPEDEDDDIE-- 136
                 ||  :  |   :: ||    : : :   ||   | |: |   |:|::::|
Db    679 EEEEEDFEEEEEDEEEYFEEEEEEEEEFEEEFE---EEEGELEEEEEEDEEEEEELEEV 735

Qy    137 -------LRGAAAAAAPPPPIEEEC--PEDLPEKFDGNPD 167
                  |     |||||       | : | |    |:
Db    736 EDLEFGTAGGEVEEGAPPPPTLPPALPPPESPPKVQPEPE 775


RESULT   5
US-09-357-014-9
; Sequence 9, Application US/09357014
; Patent No. 6291645
;  GENERAL INFORMATION:
;       APPLICANT: Jaekyoon Shin, Insil Joung, Ratna K. Vadlamudi
;                  and Jack L. Strominger
;       TITLE OF INVENTION: p62 POLYPEPTIDES, RELATED POLYPEPTIDES
;                           AND USES THEREFOR
;       NUMBER OF SEQUENCES: 22
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: LAHIVE & COCKFIELD
;            STREET: 60 State Street, Suite 510
;            CITY: Boston
;            STATE: Massachusetts
;            COUNTRY: USA
;            ZIP: 02109-1875
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
```

```
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.25
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/357,014
;             FILING DATE: 19-Jul-1999
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 08/574,959
;             FILING DATE: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Mandragouras, Amy E.
;             REGISTRATION NUMBER: 36,207
;             REFERENCE/DOCKET NUMBER: DFN-008
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (617)227-7400
;             TELEFAX: (617)227-5941
;   INFORMATION FOR SEQ ID NO: 9:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 905 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 9:
US-09-357-014-9


  Query Match             5.8%;  Score 125;  DB 4;  Length 905;
  Best Local Similarity   20.5%;  Pred. No. 0.0055;
  Matches   45;  Conservative   22;  Mismatches   75;  Indels   78;  Gaps    7;

Qy      22 GLHPHRSEATAGRSPPTPTVTLGP--------------------DCPPPPPPPPP---- 56
           || |       :| :|| |    ||                         |||||||||
Db     560 GLPPLPPPPPSGATPP-PIAPTGPPTASPPVPAKEEPEELPAAPGPLPPPPPPPPPVPGP 618

Qy      57 ------------------------------NNNNNNNSKHTGHKSACVPNMTERRRDE-- 84
                                           |:::   :  | :        |   :|
Db     619 VXLPPPQLVPEGTPGGGGPPALEEDLTVININSSDEEEEEEGEEEEEEEEEEEEEEEEE 678

Qy      85 ------LSEEINNLREKVMKQSEENNNLQSQVQKLTEENTTLREQVEPTPEDEDDDIE-- 136
                 ||  :  |   :: ||    : : :   ||   | |: |   |:|::::|
Db     679 EEEEEDFEEEEEDEEEYFEEEEEEEEEFEEEFE---EEEGELEEEEEEEDEEEEELEEV 735

Qy     137 -------LRGAAAAAAPPPPIEEEC--PEDLPEKFDGNPD 167
                   |     |||||       | : | |    |:
Db     736 EDLEFGTAGGEVEEGAPPPPTLPPALPPPESPPKVQPEPE 775


RESULT   6
US-08-574-959A-7
; Sequence 7, Application US/08574959A
; Patent No. 5962224
;  GENERAL INFORMATION:
;    APPLICANT:  Jaekyoon Shin, Insil Joung, Ratna K. Vadlamudi
;    APPLICANT:  and Jack L. Strominger
;    TITLE OF INVENTION:  p62 POLYPEPTIDES, RELATED POLYPEPTIDES
;    TITLE OF INVENTION:  AND USES THEREFOR
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LAHIVE & COCKFIELD
;      STREET:  60 State Street, Suite 510
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02109-1875
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/574,959A
;      FILING DATE:  19-DEC-95
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Mandragouras, Amy E.
;      REGISTRATION NUMBER:  36,207
;      REFERENCE/DOCKET NUMBER:  DFN-008
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)227-7400
;      TELEFAX:  (617)227-5941
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1135 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-574-959A-7


  Query Match             5.8%;  Score 125;  DB 2;  Length 1135;
  Best Local Similarity   20.5%;  Pred. No. 0.0075;
  Matches   45;  Conservative   22;  Mismatches   75;  Indels   78;  Gaps    7;

Qy      22 GLHPHRSEATAGRSPPTPTVTLGP--------------------DCPPPPPPPPP---- 56
           || |       :| :|| |    ||                         |||||||||
Db     790 GLPPLPPPPPSGATPP-PIAPTGPPTASPPVPAKEEPEELPAAPGPLPPPPPPPPPVPGP 848

Qy      57 ------------------------------NNNNNNNSKHTGHKSACVPNMTERRRDE-- 84
                                           |:::   :  | :        |   :|
Db     849 VXLPPPQLVPEGTPGGGGPPALEEDLTVININSSDEEEEEEGEEEEEEEEEEEEEEEEE 908

Qy      85 ------LSEEINNLREKVMKQSEENNNLQSQVQKLTEENTTLREQVEPTPEDEDDDIE-- 136
                 ||  :  |   :: ||    : : :   ||   | |: |   |:|::::|
Db     909 EEEEEDFEEEEEDEEEYFEEEEEEEEEFEEEFE---EEEGELEEEEEEEDEEEEELEEV 965

Qy     137 -------LRGAAAAAAPPPPIEEEC--PEDLPEKFDGNPD 167
                   |     |||||       | : | |    |:
Db     966 EDLEFGTAGGEVEEGAPPPPTLPPALPPPESPPKVQPEPE 1005


RESULT   7
US-09-357-014-7
; Sequence 7, Application US/09357014
; Patent No. 6291645
;  GENERAL INFORMATION:
;       APPLICANT: Jaekyoon Shin, Insil Joung, Ratna K. Vadlamudi
;                  and Jack L. Strominger
;       TITLE OF INVENTION: p62 POLYPEPTIDES, RELATED POLYPEPTIDES
;                           AND USES THEREFOR
;       NUMBER OF SEQUENCES: 22
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: LAHIVE & COCKFIELD
;            STREET: 60 State Street, Suite 510
;            CITY: Boston
;            STATE: Massachusetts
;            COUNTRY: USA
;            ZIP: 02109-1875
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.25
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/357,014
;            FILING DATE: 19-Jul-1999
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: 08/574,959
;            FILING DATE: <Unknown>
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Mandragouras, Amy E.
;            REGISTRATION NUMBER: 36,207
;            REFERENCE/DOCKET NUMBER: DFN-008
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (617)227-7400
;            TELEFAX: (617)227-5941
;  INFORMATION FOR SEQ ID NO: 7:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 1135 amino acids
```

```
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 7:
US-09-357-014-7


  Query Match              5.8%;  Score 125;  DB 4;  Length 1135;
  Best Local Similarity   20.5%;  Pred. No. 0.0075;
  Matches   45;  Conservative   22;  Mismatches   75;  Indels   78;  Gaps    7;

Qy      22 GLHPHRSEATAGRSPPTPTVTLGP--------------------DCPPPPPPPPP---- 56
           || |      :| :|| |    ||                     |||||||||
Db     790 GLPPLPPPPPSGATPP-PIAPTGPPTASPPVPAKEEPEELPAAPGPLPPPPPPPPPVPGP 848

Qy      57 ----------------------------NNNNNNNSKHTGHKSACVPNMTERRRDE-- 84
                                        |:::   :  | :       |   :|
Db     849 VXLPPPQLVPEGTPGGGGPPALEEDLTVININSSDEEEEEEGEEEEEEEEEEEEEEEE 908

Qy      85 ------LSEEINNLREKVMKQSEENNNLQSQVQKLTEENTTLREQVEPTPEDEDDDIE-- 136
                 ||  :  |   :: ||    : : :   ||   | |: |   |:|::::|
Db     909 EEEEEDFEEEEEDEEEYFEEEEEEEEEFEEEFE---EEEGELEEEEEEEDEEEEELEEV 965

Qy     137 -------LRGAAAAAAPPPPIEEEC--PEDLPEKFDGNPD 167
                 |   |||||     | : | |    |:
Db     966 EDLEFGTAGGEVEEGAPPPPTLPPALPPPESPPKVQPEPE 1005


RESULT   8
US-09-209-525-54
; Sequence 54, Application US/09209525
; Patent No. 6303770
; GENERAL INFORMATION:
;  APPLICANT: Lok, Si
;  APPLICANT: Conklin, Darrell C.
;  APPLICANT: Parrish, Julia E.
;  TITLE OF INVENTION: Mammalian Alpha Helical Protein-1
;  FILE REFERENCE: 97-71
;  CURRENT APPLICATION NUMBER: US/09/209,525
;  CURRENT FILING DATE: 1998-12-10
;  NUMBER OF SEQ ID NOS: 56
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 54
;   LENGTH: 80
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-209-525-54


  Query Match              5.6%;  Score 120;  DB 4;  Length 80;
  Best Local Similarity   32.9%;  Pred. No. 0.00053;
  Matches   25;  Conservative   14;  Mismatches   37;  Indels    0;  Gaps    0;

Qy     159 PEKFDGNPDMLAPFMAQCQIFMEKSTRDFSVDRVRVCFVTSMMTGRAARWASAKLERSHY 218
           || |:|    |  |: |   :|  :   |  | ::| |: |::|| |  |    :|
Db       2 PETFNGESSRLPEFIVQTASYMLVNENRFCNDAMKVAFLISLLTGEAEEWVVPYIEMDSP 61

Qy     219 LMHNYPAFMMEMKHVF 234
           :: :| ||: |||  |
Db      62 ILGDYRAFLDEMKQCF 77


RESULT   9
US-09-368-590-2
; Sequence 2, Application US/09368590
; Patent No. 6187563
; GENERAL INFORMATION:
;  APPLICANT: Solimena, Michele
;  TITLE OF INVENTION: INTERACTING POLYPEPTIDES FOR
;  TITLE OF INVENTION: AUTOANTIGENS OF AUTOIMMUNE DISEASES
;  FILE REFERENCE: 101918-200 (OCR-941)
;  CURRENT APPLICATION NUMBER: US/09/368,590
;  CURRENT FILING DATE: 1999-08-04
;  EARLIER APPLICATION NUMBER: 60/095,657
;  EARLIER FILING DATE: 1998-08-07
;  NUMBER OF SEQ ID NOS: 8
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2
;   LENGTH: 2293
;   TYPE: PRT
;   ORGANISM: Human
US-09-368-590-2


  Query Match              5.5%;  Score 118.5;  DB 4;  Length 2293;
  Best Local Similarity   23.3%;  Pred. No. 0.079;
  Matches   91;  Conservative   41;  Mismatches  145;  Indels  113;  Gaps   20;

Qy      29 EATAGRSPPTPTV---------TLGPDCPPPPPPPPPNNNNNNNSKHTGHKSACVPNMTE 79
           :|   : |||| :          |     |   |             |::    | :
Db    1846 KAEQSKQPPTPLLGRKFFGDPTELAAKAAPLLRP------------GGYERGLEP-LAR 1891

Qy      80 RRRDELSEEINNLREKVMKQSEENNNLQSQVQKLTEENTTLREQVEP-----TPEDEDDD 134
           |  | || |:   |   ::|  :   || :: :| |    :  :|||      |||  :
Db    1892 RASDTLSAEVRT-RVGYVRQELKPERLQPRIDRLPE----IPGRVEPAALPAAPEDAAET 1946

Qy     135 IELRGAAAAAAPPPPIEEEC--PEDLPEKFDGNPDMLAPFMAQCQIFMEKSTRDFSVDRV 192
                ||   | |  :|     |:|| :                    :  |  |||:
Db    1947 PATPAAAEQVRPRPERQESADRAEELPRR------------------RRPERQESVDQS 1987

Qy     193 RVCFVTSMMTGRAARWASAKLERSHYLMHNYPAFMMEMKHVFEDPQRREVAKRKIRRLRQ 252
                :    |  |  ||: | :| |              :|  :||   :|: ||| :
Db    1988 E----EAARRRPERQESAEHEAAHSLTLGR----------YEQMERRR--ERRERRLER 2031

Qy     253 GMGSVIDYSNAFQMIAQDLDWNEPALIDQYHEGLSDHIQEELSHLEVAKSLSALIGQCIH 312
                : |     |  ||   :  | |          | |:|   |    | |  |  :
Db    2032 Q-----ESSEQEMPIRGDLVKGKATLAD---------IVEQLQEKEAGPGLPA--GPSLP 2075

Qy     313 IERRLARAAAARKPRSPPRALVLPHIASHHQVDPTEPVGGARMRLTQEEKERRRKLNLCL 372
             | |         |   |  | ||          |  |   || |     : | |:
Db    2076 QPREL-------PPGRLPNGLELPERTPR----PDRP--RARDRPKPRRRPRPRE----- 2117

Qy     373 YCGTGGHYADNCPAKASKSSPA--GNSPAP 400
             | ||        :: |:|:||  |::|||
Db    2118 -GGEGG------GSRRSRSAPAQGGSAPAP 2140


RESULT  10
US-08-539-205A-6
; Sequence 6, Application US/08539205A
; Patent No. 6001619
;  GENERAL INFORMATION:
;    APPLICANT:  Beach, David H.
;    APPLICANT:  Caligiuri, Maureen
;    APPLICANT:  Nefsky, Bradley
;    TITLE OF INVENTION:  Ubiquitin Ligases, and Uses Related Thereto
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FOLEY, HOAG & ELIOT LLP
;      STREET:  One Post Office Square
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109-2170
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/539,205A
;      FILING DATE:  04-OCT-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Vincent, Matthew P.
;      REGISTRATION NUMBER:  36,709
;      REFERENCE/DOCKET NUMBER:  CSV-005.01
```

```
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 832-1000
;      TELEFAX:  (617) 832-7000
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  834 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-539-205A-6


  Query Match             5.5%;  Score 118;  DB 3;  Length 834;
  Best Local Similarity   22.2%;  Pred. No. 0.021;
  Matches   50;  Conservative   29;  Mismatches   66;  Indels   80;  Gaps   12;

Qy      50 PPPPPPP--------------NNNNNNNSKHTGHKSACVPNMTERRRDELSEEINNLRE 94
           |||| ||               |:||     |       |::     |  ||  ||:|:
Db      70 PPPPLPPGWEEKVDNLGRTYYVNHNNRTTQWHR-------PSL----MDVSSESDNNIRQ 118

Qy      95 KVMKQSEENNNLQSQVQKLTEENTTLREQVEPTPEDEDD----------DIELRG-AAAA 143
             : |   :   :|:          : | :|| | :  |          :: : | :
Db     119 --INQEAAHRRFRSR--------RHISEDLEPEPSEGGDVPEPWETISEEVNIAGDSLGL 168

Qy     144 AAPPPPIE-------EECPEDLPEKFDGNPDMLAPFMAQCQIFMEKSTRDFSVDRVRVCF 196
           | ||||:        :|  |:|  :    ||      :  : |     |: |  |:| |
Db     169 ALPPPPVSPGSRTSPQELSEELSRRLQITPD------SNGEQFSSLIQREPS-SRLRSCS 221

Qy     197 VTSMMT----------------GRAARWASAK--LERSHYLMHN 222
           ||  :                |  : |   |    |::|: ||
Db     222 VTDAVAEQGHLPPPSVAYVHTTPGLPSGWEERKDAKGRTYYVNHN 266


RESULT  11
US-09-540-245A-18
; Sequence 18, Application US/09540245A
; Patent No. 6270984
; GENERAL INFORMATION:
;  APPLICANT:  Goodman, Corey
;  APPLICANT:  Kid, Thomas
;  APPLICANT:  Brose, Katja
;  APPLICANT:  Tessier-Lavigne, Marc
;  TITLE OF INVENTION:  Modulating Robo: Ligand Interactions
;  FILE REFERENCE:  B98-031-3
;  CURRENT APPLICATION NUMBER: US/09/540,245A
;  CURRENT FILING DATE:  2000-03-31
;  PRIOR APPLICATION NUMBER:  60/065,544
;  PRIOR FILING DATE:  1997-11-14
;  PRIOR APPLICATION NUMBER:  60/081,057
;  PRIOR FILING DATE:  1998-04-07
;  NUMBER OF SEQ ID NOS:  20
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 18
;   LENGTH: 1651
;   TYPE: PRT
;   ORGANISM: human
US-09-540-245A-18


  Query Match             5.4%;  Score 115;  DB 4;  Length 1651;
  Best Local Similarity   19.0%;  Pred. No. 0.1;
  Matches   81;  Conservative   47;  Mismatches  171;  Indels  128;  Gaps   15;

Qy      14 SGRGGQDPGLHPHRSEATAGRSPPTPTVTLGPDCPPPPPPPPPNNN----NNNNSKHTGH 69
           | ||    |   |:  |   : |    :      ||||  |||::|    | :  :
Db    1153 SDRGSSTSGSQGHKKGARTPKVPKQGGMNWADLLPPPPAHPPPHSNSEEYNISVDESYDQ 1212

Qy      70 KSAC--VPNMTERRRDELSEEINN------LREKVMKQSEENNNLQSQVQKLTEENTTLR 121
           :  |   |     ::||| || :       :|     :  : : ||          |:
Db    1213 EMPCPVPPARMYLQQDELEEEEDERGPTPPVRGAASSPAAVSYSHQSTATLTPSPQEELQ 1272

Qy     122 EQVEPTPEDE---DDDIELRGAAAAAAPPP-------------PIEEECPEDLPEKFDG 164
             ::  ||:       : |    :  |||              |:  :   | ||: :
Db    1273 PMLQDCPEETGHMQHQPDRRRQPVSPPPPPRPISPPHTYGYISGPLVSDMDTDAPEEEED 1332

Qy     165 NPDMLAPFMAQCQIFME--KSTRDFSVDRVRVCFVTSMMTGRAARWASAKLERSHYLMHN 222
             ||    |   :: :   : |   ||  :      ||: |    | ||  | :
Db    1333 EADMEVAKMQTRRLLLRGLEQTPASSVGDLESSVTGSMING----WGSASEEDN------ 1382

Qy     223 YPAFMMEMKHVFEDPQRREVAKRKIRRLRQGMGSVIDYSNAFQMIAQDLDWNEPALIDQY 282
                                       :  |  ||     :|     | |: :
Db    1383 --------------------------ISSGRSSVSSSDGSF---FTDADFAQAVAAAAE 1412

Qy     283 HEGLSDHIQEELSHLEVAKSLSALIGQCIHIERRLARAAAARKPRSPPRALVLPHIASHH 342
           : ||                        : ||  : || |:             |  :
Db    1413 YAGLK------------------------VARRQMQDAAGRR-----------HFHASQ 1436

Qy     343 QVDPTEPV------GGARMRLTQEEKERRRKLNLCLYCGTGGH-----YADNCPAKASKS 391
             || ||        | |: |:  |: : :          ||     | |: |
Db    1437 CPRPTSPVSTDSNMSAAVMQKTRPAKKLKHQ---------PGHLRRETYTDDLP-PPPVP 1486

Qy     392 SPAGNSP 398
           ||  ||
Db    1487 PPAIKSP 1493


RESULT  12
US-09-026-343-2
; Sequence 2, Application US/09026343
; Patent No. 6008018
;  GENERAL INFORMATION:
;    APPLICANT:  DUAN, D. ROXANNE
;    APPLICANT:  SHILATIFARD, ALI
;    APPLICANT:  CONAWAY, JOAN W.
;    APPLICANT:  CONAWAY, RONALD C.
;    TITLE OF INVENTION:  ELL2, A New Member of an ELL Family of
;    TITLE OF INVENTION:  RNA Polymerase II Elongation Factors
;    NUMBER OF SEQUENCES:  34
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
;      STREET:  1100 NEW YORK AVENUE, SUITE 600
;      CITY:  WASHINGTON
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20005-3934
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/026,343
;      FILING DATE:  Herewith
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/038,447
;      FILING DATE:  19-FEB-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  GOLDSTEIN, JORGE A.
;      REGISTRATION NUMBER:  29,021
;      REFERENCE/DOCKET NUMBER:  1488.0880001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 371-2600
;      TELEFAX:  (202) 372-2540
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  640 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-026-343-2


  Query Match             5.3%;  Score 114;  DB 3;  Length 640;
  Best Local Similarity   19.4%;  Pred. No. 0.034;
  Matches   64;  Conservative   55;  Mismatches  125;  Indels   86;  Gaps   13;
```

```
Qy      21 PGLHPH---RSEATAGRSPPTPTVTLGPDCPPPP--------PPPPPNNNNNNNSKHTGH 69
           | |: |    || :|   |  |      |  || |        ||   |:|:|: |   |
Db     346 PTLNGHLNPTSEKSAAGLPLPPAAAAIPTPPPLPSTYLPISHPPQIVNSNSNSPSTPEGR 405

Qy      70 KSACVPNMTERRRDELSEE--------------INNLREKVMKQSEENNNLQSQVQKLT 114
            :  :|  :  : | : |:                ::  |  |  |||:::  :  |
Db     406 GTQDLPVDSFSQNDSIYEDQQDKYTSRTSLETLPPGSVLLKCPKPMEENHSMSHKKSKKK 465

Qy     115 EENTTLREQVE----PTPEDEDDDIELRGAAAAAAPPPP-----IEEECPEDLPEKFDGN 165
           :    ::|::     | |::::|::     |       |     ::|:|   :
Db     466 SKKHKEKDQIKKHDIETIEEKEEDLKREEEIAKLNNSSPNSSGGVKEDCTASMEPSAIEL 525

Qy     166 PDMLAPFMAQCQIFMEKSTR-DFSVDRVRVCFVTSMMTGRAARWASAKLERSHYLMHNYP 224
           || |  ::|       :: : ||: :                                  |
Db     526 PDYLIKYIAIVSYEQRQNYKDDFNAE----------------------------YDEYR 556

Qy     225 AFMMEMKHVFEDPQRREVAKRKIRRLRQGMGSVIDYSNAFQMIAQD---LDWNEPALIDQ 281
           |    |: |      |   |   :| |   ||  :| |  : : |:   :  : |
Db     557 ALHARMETV-----ARRFIKLDAQRKRLSPGSK-EYQNVHEEVLQEYQKIKQSSP----N 606

Qy     282 YHEG--LSDHIQEELSHLEVAKSLSALIGQ 309
           |||     :::  :|:|::       |||:
Db     607 YHEEKYRCEYLHNKLAHIK------RLIGE 630


RESULT 13
US-08-545-860D-48
; Sequence 48, Application US/08545860D
; Patent No. 6040140
;  GENERAL INFORMATION:
;    APPLICANT:  Croce, Carlo
;    APPLICANT:  Canaani, Eli
;    TITLE OF INVENTION:  Diagnostics, Therapeutics and Methods
;    TITLE OF INVENTION:  for Detection and Treatment of Acute Leukemias
;    TITLE OF INVENTION:  Resulting from Chromosome Abnormalities in the All-1 Region
;    NUMBER OF SEQUENCES:  94
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Woodcock, Washburn, Kurtz, Mackiewitz &
;      ADDRESSEE:  No. 6040140ris
;      STREET:  One Liberty Place, 46th floor
;      CITY:  Philadelphia
;      STATE:  Pennsylvania
;      COUNTRY:  USA
;      ZIP:  19103
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/545,860D
;      FILING DATE:  07-MAR-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US94/04496
;      FILING DATE:  22-APR-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US92/10930
;      FILING DATE:  09-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/327,392
;      FILING DATE:  19-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/320,559
;      FILING DATE:  11-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/062,443
;      FILING DATE:  14-MAY-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/971,094
;      FILING DATE:  30-OCT-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/888,839
;      FILING DATE:  27-MAY-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/805,093
;      FILING DATE:  11-DEC-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  DeLuca Esq., Mark
;      REGISTRATION NUMBER:  33,229
;      REFERENCE/DOCKET NUMBER:  TJU-1262
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (215) 568-3100
;      TELEFAX:  (215) 568-3439
;  INFORMATION FOR SEQ ID NO:  48:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1612 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-545-860D-48


  Query Match             5.2%;  Score 112;  DB 3;  Length 1612;
  Best Local Similarity   19.7%;  Pred. No. 0.19;
  Matches   60;  Conservative   41;  Mismatches  138;  Indels   66;  Gaps    9;

Qy      31 TAGRSPPTPTVTLGP---DCPPPPPPPPPNNNNN------------NNSKHTGHKSACV 74
           |    | ||      |   | |||||||| :   :              ::   |  || |
Db    1329 TPAAIPATPVAVSQPIRTDLPPPPPPPPVHYAGDFDGMSMDLPLPPPPSANQIGLPSAQV 1388

Qy      75 PNMTERRRDELSEEINNLREKVMKQSEENNNLQSQVQKLTEENTTLREQVEPTPEDEDDD 134
                |:|:|        :||   : |     : | :|| :  |   : : | |
Db    1389 AAAERRKREEHQRWYE--KEKAPLEEERERKRREQERKLGQMRT---QSLNPAPFSPLTA 1443

Qy     135 IELRGAAAAAAPPPP---IEEECPEDLPEKFDGNPDMLAPFMAQCQIFMEKSTRDFSVDR 191
           :::    :    |    | |  |: |   :                    :  :  :| :
Db    1444 QQMKPEKPSTLQRPQETVIRELQPQQQPRTIE----------------RRDLQYITVSK 1486

Qy     192 VRVCFVTSMMTGRAARWASAKLERSHYLMHNYPAFMMEMKHVFEDPQRREVAKRKIRRLR 251
             :    |:      | |  |||:    ||       |:: :   | |      ::|:|
Db    1487 EELSSGDSLSPDPWKRDAKEKLEKQQQ-MHIVDMLSKEIQELQSKPDRSAEESDRLRKLM 1545

Qy     252 QGMGSVIDYSNAFQMIAQDLDWNEPALIDQYHEGLSDHIQEELSHLEVAKSLSALIGQCI 311
                              |:|     : :  :   |  :||   ::     : || | :
Db    1546 ------------------LEWQFQKRLQESKQKDEDDEEEEDDDVD-----TMLIMQRL 1581

Qy     312 HIERR 316
           |||
Db    1582 EAERR 1586


RESULT 14
PCT-US94-04496-48
; Sequence 48, Application PC/TUS9404496
;  GENERAL INFORMATION:
;    APPLICANT:  Croce, Carlo
;    APPLICANT:  Canaani, Eli
;    TITLE OF INVENTION:  Diagnostics, Therapeutics and Methods
;    TITLE OF INVENTION:  for Detection and Treatment of Acute Leukemias
;    TITLE OF INVENTION:  Resulting from Chromosome Abnormalities in the All-1
;    NUMBER OF SEQUENCES:  86
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Woodcock, Washburn, Kurtz, Mackiewitz &
;      ADDRESSEE:  Norris
;      STREET:  One Liberty Place, 46th floor
;      CITY:  Philadelphia
;      STATE:  Pennsylvania
;      COUNTRY:  USA
;      ZIP:  19103
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
```

```
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US94/04496
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  DeLuca Esq., Mark
;      REGISTRATION NUMBER:  33,229
;      REFERENCE/DOCKET NUMBER:  TJU-1242
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (215) 568-3100
;      TELEFAX:  (215) 568-3439
;  INFORMATION FOR SEQ ID NO:  48:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1612 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
PCT-US94-04496-48


  Query Match             5.2%;  Score 112;  DB 5;  Length 1612;
  Best Local Similarity   19.7%;  Pred. No. 0.19;
  Matches   60;  Conservative   41;  Mismatches  138;  Indels   66;  Gaps    9;

Qy      31 TAGRSPPTPTVTLGP---DCPPPPPPPPPNNNNN----------NNSKHTGHKSACV 74
           |    | ||      |   | ||||||| :   :              ::   |  || |
Db    1329 TPAAIPATPVAVSQPIRTDLPPPPPPPPVHYAGDFDGMSMDLPLPPPPSANQIGLPSAQV 1388

Qy      75 PNMTERRRDELSEEINNLREKVMKQSEENNNLQSQVQKLTEENTTLREQVEPTPEDEDDD 134
             |:|:|       :||   : |     : | :|| :  |   : : | |
Db    1389 AAAERRKREEHQRWYE--KEKAPLEEERERKRREQERKLGQMRT---QSLNPAPFSPLTA 1443

Qy     135 IELRGAAAAAAPPPP---IEEECPEDLPEKFDGNPDMLAPFMAQCQIFMEKSTRDFSVDR 191
           :::    :    |    | |  |: |   :                  :  :  :| :
Db    1444 QQMKPEKPSTLQRPQETVIRELQPQQQPRTIE----------------RRDLQYITVSK 1486

Qy     192 VRVCFVTSMMTGRAARWASAKLERSHYLMHNYPAFMMEMKHVFEDPQRREVAKRKIRRLR 251
             :    |:      | |  |||:    ||       |:: :   | |      ::|:|
Db    1487 EELSSGDSLSPDPWKRDAKEKLEKQQQ-MHIVDMLSKEIQELQSKPDRSAEESDRLRKLM 1545

Qy     252 QGMGSVIDYSNAFQMIAQDLDWNEPALIDQYHEGLSDHIQEELSHLEVAKSLSALIGQCI 311
                          |:|    : :  :   |  :||   ::     : || | :
Db    1546 ------------------LEWQFQKRLQESKQKDEDDEEEEDDDVD-----TMLIMQRL 1581

Qy     312 HIERR 316
             |||
Db    1582 EAERR 1586


RESULT  15
US-09-370-368-9
; Sequence 9, Application US/09370368
; Patent No. 6258932
; GENERAL INFORMATION:
;  APPLICANT: Anders Vahlne
;  TITLE OF INVENTION: PEPTIDES THAT BLOCK VIRAL INFECTIVITY
;  TITLE OF INVENTION: AND METHODS OF USE THEREOF
;  FILE REFERENCE: TRIPEP.003A
;  CURRENT APPLICATION NUMBER: US/09/370,368
;  CURRENT FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 10
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 9
;   LENGTH: 538
;   TYPE: PRT
;   ORGANISM: Moloney Murine Leukemia Virus
US-09-370-368-9


  Query Match             5.2%;  Score 110.5;  DB 4;  Length 538;
  Best Local Similarity   19.0%;  Pred. No. 0.057;
  Matches   96;  Conservative   51;  Mismatches  184;  Indels  175;  Gaps   22;

Qy      21 PGLHPHRS--------EATAGRSPPTPTVTLGPDCPPPPPPP----------PPNNNNN 61
           || | |          || |   ||    . : |   | ||||          || :
Db      71 PGPHGHPDQVPYIVTWEALAFDPPP----WVKPFVHPKPPPPLLPSAPSLPLEPPLS--- 123

Qy      62 NNSKHTGHKSACVPNMTERRRDELSEEINNLREKVMKQSEENNNLQSQVQKLTEENTTLR 121
                 |  :|:  | :|              : |    |:    |   :  |||:    |
Db     124 -----TPPQSSLYPALTPSL---------GAKPKPQVLSDSGGPL---IDLLTEDPPPYR 166

Qy     122 EQVEPTPEDEDDDI-ELRGAAAAAAP------------PPPIEEECPEDLPEKFDGNPDM 168
           :   | | | | |   |   |  | |             ||  :    :  | :  ||  :
Db     167 DP-RPPPSDRDGDSGEATPAGEAPDPSPMASRLRGRREPPVADSTTSQAFPLRTGGNGQL 225

Qy     169 -LAPFMAQCQIFMEKSTRDFSVDRVRV-CFVTSMMTGRAARW------------------ 208
            || :      : :   || |  ::   : |::      |
Db     226 QYWPFSSSDLYNWKNNNPSFSEDPGKLTALIESVLITHQPTWDDCQQLLGTLLTGEEKQR 285

Qy     209 ------------------------ASAKLER----------SHYLMH------------ 221
                                    |:  |||          ::|:|
Db     286 VLLEARKAVRGDDGRPTQLPNEVDAAFPLERPDWEYTTQAGRNHLVHYRQLLIAGLQNAG 345

Qy     222 NYPAFMMEMKHVFEDPQRREVAKRKIRRLRQGMGSVIDYS----NAFQMIAQDLDWNEPA 277
             |  : ::| : : |   |     : ||::       |          ::    |
Db     346 RSPTNLAKVKGITQGP--NESPSAFLERLKEAYRRYTPYDPEDPGQETNVSMSFIWQSAP 403

Qy     278 LIDQYHEGLSDHIQEELSHLEVAKSLSALIGQCIHI----------------------ER 315
           | : | |     |:| :   |:|  |: :  |                          ||
Db     404 DIGRKLERL-----EDLRN----KTLGDLVREAERIFNKRETPEEREERIRREREEKEER 454

Qy     316 RLARAAAARKPRSPPRALVLPHIASHHQVDPTEPVGGARMRLTQEEKERRRK---LNLCL 372
           |        | |   |          | ::      :  |  ::| ||||     : |
Db     455 RRTEDEQKEKERDRRR---------HREMSRLLATVVSGQRQDRQEGERRRSQLDCDQCT 505

Qy     373 YCGTGGHYADNCPAKASKSSPAGNSP 398
           ||   ||:| :|| :     | |  |
Db     506 YCEEQGHWAKDCPRR--PRGPRGPRP 529


Search completed: October  4, 2002, 15:38:43
Job time: 25 sec
```